

AN INVESTIGATION OF THE RELATIONSHIP BETWEEN BEHAVIORAL FEEDBACK
ON SOCIAL-EMOTIONAL SKILLS ON REPORT CARDS AND ACADEMIC

ACHIEVEMENT

SUBMITTED TO THE FACULTY

OF

THE GRADUATE SCHOOL OF APPLIED AND PROFESSIONAL PSYCHOLOGY

OF

RUTGERS,

THE STATE UNIVERSITY OF NEW JERSEY

BY

JOSEPH FERRITO

IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE

OF

DOCTOR OF PSYCHOLOGY

NEW BRUNSWICK, NEW JERSEY

OCTOBER 2015

APPROVED:

Maurice J. Elias, PhD

Anne Gregory, PhD

DEAN:

Stanley Messer, PhD.

Copyright 2015 by Joseph Ferrito

Abstract

Instruction and emphasis on Social-Emotional Learning (SEL) have increasingly become a focus of educational policy at the local, state and federal levels. As schools are required to emphasize SEL skills, accurate and feasible methods to assess student progress on skill development are necessary. Systematic reviews of SEL measures have identified few that are psychometrically sound and valid in assessing essential SEL skill areas. Of these measures, even fewer demonstrate feasibility and scalability for use in schools across the nation or abroad. This study seeks to investigate the potential utility of measuring SEL via a behavior feedback mechanism already in place in schools; report card comments. Preliminary research by the SEL Lab at Rutgers University has identified report card comments to be partially representative of SEL skills and to have a relationship with achievement. The current sample of 113 students enrolled in a suburban elementary school received a report card that included a typical comment menu and Likert ratings items based on SEL programming in practice. Results indicated typical comment and Likert rating items were partially representative of SEL skills. Both comments and explicit SEL Likert rating items were found to be associated with achievement. However, after controlling for previous achievement, only SEL Likert rating items were found to have a significant link to achievement. These findings show that an alternative report card comment structure that included Likert rating items explicitly related to SEL demonstrated stronger concurrent validity, accounted for a more significant portion of unique variance, beyond typical comments. The school developed and effectively implemented the adapted report card comments section in this study, suggesting its feasibility and scalability as a method of SEL assessment in schools.

Acknowledgements

This project would not have occurred without support and encouragement from educators, fellow researchers, and family. The tireless efforts of the research team at the SEL Lab at Rutgers University made this and many other of my research endeavors possible. Dr. Maurice Elias, the director of that lab, has served not only as my dissertation chair, but as a mentor and perpetual source of inspiration, motivation and support. Dr. Anne Gregory, my second chair, has served as a model of passion, rigor, and dedication through her exemplary research and captivating instructional style. I feel beyond lucky to have encountered both Maurice and Anne early in my time at GSAPP. I thank you both sincerely not only for your support on this project, but for the impact you continue to have in shaping my identity as a psychologist.

I feel remarkably fortunate in knowing that none of this would have been possible without the love and support of my incredible family. My parents, Rose and Al, continue to redefine the numerous ways in which we as people can positively impact and connect with those around us. My sister, Laurel, through her compassion and understanding, has been my role model in entering the field of psychology. My brother, Vince, has demonstrated how engaging with drive and dedication can lead you beyond your expectations. Lastly, I would like to thank my beautiful girlfriend, Lindsey. Your unwavering support of my professional goals has led us to travel thousands of miles together and leave much of what we have ever known. The fact that you perpetually offer support through brilliant insight and thoughtfulness without hesitation is remarkable.

Table of Contents

	PAGE
ABSTRACT	ii
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	vii
CHAPTER	
I. INTRODUCTION	1
SEL as a Form of Prevention and Promotion?.....	2
Policy Regarding SEL.....	5
The Need for SEL Assessment	7
Current State of SEL Assessment	8
The Multifaceted Nature of SEL Skills	9
Psychometric Properties.....	11
Scope, Specificity, Purpose, and Setting	12
Types of Measures	14
Performance-Based Assessments.....	15
Student Self-Report.....	16
Observer Report.....	17
Multiple Informants	19
Recommendations for Selecting SEL Measures	20
Measuring SEL Competencies on 'The Other Side of the Report Card'	21

The Current Study.....	27
II. METHOD	30
Setting and Participants.....	30
Measures	31
Report Card Variables.....	31
Non-Report Card Variables	32
Effect Sizes	32
III. RESULTS	34
RQ1 Face and Content Validity.....	34
Qualitative Categorization	34
Exploratory Factor Analysis	36
Data Reduction	37
RQ2 Concurrent Validity.....	39
H2a: Positive Comments and Higher Ratings Associated with Higher GPA	39
H2a: Positive Comments and Higher Ratings Associated with Higher NJASK.....	40
H3a: Comments and Ratings Related to Social-Awareness and Relationship Skills Associated with Achievement	41
H3a: Accounting for Previous Achievement in Associations Between Comments, Ratings, and Achievement.....	42
IV. DISCUSSION.....	45
Comparison with Existing Literature.....	45
Considerations and Potential Explanations for Current Findings.....	49
Limitations and Suggestions for Future Research	55

Implications and Suggestions for Current Practice.....61

REFERENCES63

APPENDIX A: CASEL 5: Definitions and Skills Examples99

APPENDIX B Copy of a De-Identified Report Card104

List of Tables

Table # 1 Definitions for the CASEL 5 Core SEL Competencies.....	71
Table # 2 Student Demographic Descriptive Statistics	72
Table # 3 Typical Report Card List of 24 Comment Options.....	73
Table # 4 "Personal and Social Development" Rating Items.....	74
Table # 5 Data Analytic Techniques for the Current Study.....	75
Table # 6 Conceptual Categorization of the 24 Typical Comments	76
Table # 7 Conceptual Categorizations of the "Personal and Social Development" Items.....	77
Table # 8 Descriptive Statistics for Cumulative Use of Comments across Marking Periods and Subject Areas	78
Table # 9 Exploratory Factor Analysis Component Loadings for Typical Comment Total Scores from All Marking Periods and Subject Areas Domains Areas	79
Table # 10 Exploratory Factor Analysis Component Loadings for "Personal and Social Development" Total Scores from All Marking Periods.....	80
Table # 11 Correlations Matrix for Variables Considered in Data Reduction Strategy	81
Table # 12 Comparisons of Achievement Means for Male and Female Groups.....	82
Table # 13 Descriptive Statistics for Achievement and Comment Variables.....	83
Table # 14 Hierarchical Regression with Cumulative GPA as the Dependent Variable.....	84
Table # 15 Hierarchical Regression with Cumulative GPA for Reading and Writing as the Dependent Variable	85
Table # 16 Hierarchical Regression with Cumulative GPA for Science and Math as the Dependent Variable	86
Table # 17 Hierarchical Regression with NJASK Language Arts Scores the Dependent Variable	87
Table # 18 Hierarchical Regression with NJASK Math Scores as the Dependent Variable	88
Table # 19 Hierarchical Regression with Cumulative GPA for Reading and Writing as the Dependent Variable and PSD Proficiency then Deficiency Total Scores	89

Table # 20 Hierarchical Regression with Cumulative GPA for Science and Math as the Dependent Variable and PSD Proficiency then Deficiency Total Scores90
Table # 21 Hierarchical Regression with Cumulative GPA for Reading and Writing as the Dependent Variable and PSD Deficiency the ProficiencyTotal Scores91
Table # 22 Hierarchical Regression with Cumulative GPA for Science and Math as the Dependent Variable and PSD Deficiency then Proficiency Total Scores92
Table # 23 Hierarchical Regression with NJASK Language Arts Scores as the Dependent Variable and PSD Proficiency then Deficiency Total Scores.....	.93
Table # 24 Hierarchical Regression with NJASK Math Scores as the Dependent Variable and PSD Proficiency then Deficiency Total Scores94
Table # 25 Hierarchical Regression with NJASK Language Arts Scores as the Dependent Variable and PSD Deficiency then ProficiencyTotal Scores.....	.95
Table # 26 Hierarchical Regression with NJASK Math Scores as the Dependent Variable and PSD Deficiency then Proficiency Total Scores96
Table # 27 Hierarchical Regression with Combined GPA for Reading and Writing for Marking Period Four as the Dependent Variable while Controlling for GPA from Marking Period One.....	.97
Table # 28 Hierarchical Regression with Combined GPA for Science and Math for Marking Period Four as the Dependent Variable while Controlling for GPA from Marking Period One.....	.98

Chapter I

Introduction

In the United States, 43.7 million or 18.6% of the adult population have a diagnosable mental illness with 9.6 million or 4.1% having a serious mental illness (SMI). In 2008, 13.4% of adults received treatment for a mental health problem and only 58.7% of those with a SMI received treatment. Similarly, approximately 20%, or one in five children, currently or previously experienced a seriously debilitating mental health disorder (NIMH, 2012). In 2014, the Center for Disease Control found that 29.9% of high school aged students felt sad or hopeless for two or more weeks. Only 50.6% of children with mental health disorders received treatment with 12 to 15 year olds being 90% more likely to receive treatment than eight to eleven-year-olds (NIMH, 2012). Estimated costs associated with the current population of youth with mental health disorders alone are variable, but include figures such as \$247 billion across the life-span (O'Connel et al., 2009). This figure includes costs of mental health services, legal involvement, child protective services, and so on. Beyond financial implications, these disorders can result in increased suffering among individuals and families, limitations to an individual's ability to achieve social and educational goals, and an increased risk for future psychopathology and suboptimal functioning (O'Connel et al., 2009).

There is a clear discrepancy between the number of adults, adolescents, and children suffering from a mental health disorder and those receiving services. All who do not receive services must rely on their own resources for support in managing symptoms that interfere significantly with their daily functioning, which may lead to later substantial financial implications for society. Being unable to meet the staggering level of need has progressively led researchers, practitioners, and policy-makers alike to focus on developing strategies to better

address the mental health of the population. Several decades of research on the course, prevalence, and treatment of mental health disorders has highlighted the efficacy of prevention and promotion based strategies in delaying or preventing the onset of such disorders; particularly when young people and early intervention are the focus, as approximately half of adult mental health disorders begin in childhood (O'Connel et al., 2009). Prevention refers to the reduction of maladaptive behaviors while promotion entails systematically increasing positive and adaptive behaviors (O'Connel et al., 2009).

The prevention of mental, emotional, and behavioral disorders (MEBD) has been advocated to occur as a three tiered approach. The first tier includes universal interventions applied to all youth. Selective interventions serve at the second tier to target youth at-risk, and the third tier entails indicated interventions for youth at high-risk of developing or exacerbating a MEBD. The common goals across these levels of intervention include addressing both risk and protective factors linked to determinants of MEBDs, such as poverty, family functioning, early childhood experiences, social skills, substance use, depression, and significantly maladaptive, aggressive, or deviant behaviors (O'Connel et al., 2009).

Social-Emotional Learning (SEL) as a Form of Prevention and Promotion

Educational systems are often targeted as settings in which prevention and promotion strategies can be implemented for a variety of reasons including the large amount of time students spend in school, the frequency of interaction with adults and peers, and the nature of the setting as one that facilitates student learning and growth. In addition, educators, policy makers, and the public largely agree that students graduating from various educational systems should be proficient in core academic subjects, able to work well with others of diverse backgrounds, practice healthy behaviors, and behave respectfully and responsibly (Association for Supervision

and Curriculum Development, 2007). In order for these goals to be met, prevention and promotion based strategies are needed.

Social-Emotional Learning (SEL) has emerged in both United States and international literature as a promising universal intervention to promote success in school and life (Durlak et al., 2011; OECD, 2015 as cited in Miyamoto, Huera, Kubacka, 2015). Rigorous developmental research has formed a definition of SEL as, “the process through which children and adults acquire and effectively apply the knowledge, attitudes, and skills necessary to understand and manage emotions, set and achieve positive goals, feel and show empathy for others, establish and maintain positive relationships, and make responsible decisions” (CASEL, 2015c). SEL includes five interrelated sets of cognitive, affective, and behavioral competencies, each of which are broken down into specific, observable behaviors. The five core SEL competencies are Self-Awareness, Self-Management, Social Awareness, Relationship Skills, and Responsible Decision-Making, and are commonly referred to as the CASEL 5. Table 1 provides brief definitions with specific skills in each of the CASEL 5 skill areas (CASEL, 2015b).

Durlak and colleagues (2011) conducted a meta-analysis of 213 school-based, universal SEL programs involving over 270,000 students in kindergarten through high school in both urban and rural settings. When compared to control groups, students at every grade level who participated in SEL programs demonstrated significantly improved social-emotional skills, attitudes (e.g., toward self, others, and school), behaviors (e.g., increased pro-social behaviors, reduced conduct issues) and academic performance (i.e., 11-percentile point improvement on achievement tests and grades). Further, this meta-analysis demonstrated that findings occurred most strongly when school staff effectively conducted SEL programs over more than one year. Among the implications of this are that these interventions are best incorporated into routine

educational practices and do not require outside personnel for effective delivery. In a similar meta-analysis, Gravesteyn (2012) examined 75 studies conducted between 1995 and 2008 on the effects of universal, school-based SEL programs. Results indicated that positive effects were present for the enhancement of social-emotional skills, positive self-image, prosocial behavior, and academic achievement, as well reduction or prevention of antisocial behavior and mental health problems and disorders.

It is important to note, however, that the CASEL 5 emerged from a diverse body of research that encompassed a variety of terminology, definitions, and conceptualizations of factors beyond intelligence that comprise a successful, contributing citizen. In recent years, the question of which core skills, competencies, and traits would best serve youth in becoming successful in school and life has driven efforts of CASEL and similar organizations to consider which of these competencies are most essential to integrate into education. Character Development is another prominent grouping of skills and traits that has gained significant support in research for promoting academic achievement and other positive outcomes for students, such as engagement in school, social skills, decreased anxiety and depression, and life satisfaction (Linkins, Niemiec, Gillham, & Mayerson, 2014). Peterson & Seligman (2004) identify 24 Values in Action (VIA) strengths that have stimulated a growing body of literature. Organizations such as Knowledge is Power Program (KIPP) Academies, a network of charter schools in New York City (NYC) with open enrollment based on a lottery system targeting minority and low-income students, have integrated character strengths into every aspect of the school day (KIPPNYC.org). Based on the work of Drs. Martin Seligman, Chris Peterson, and Angela Duckworth, KIPPNYC has selected seven character strengths to emphasize (i.e., Grit, Zest, Self-control, Optimism, Gratitude, Social Intelligence, & Curiosity) and attribute much of

student success to this emphasis (e.g., 94% of KIPP NYC alumni graduated high school, nearly double the rate for minority students in NYC; 40% of KIPP NYC alumni earned a B.A. degree, compared to 10% of low-income students nationally) (KIPPNYC.org).

As focus on areas such as SEL and Character continues to expand, multiple organizations have reviewed literature across the related fields of study to identify groupings of essential skills, competencies, and traits with the strongest evidence for promoting academic and life success. For example, The Tauck Family Foundation in collaboration with Child Trends conducted a recent review of literature (Chien, Harbin, Goldhagen, Lippman, & Walker, 2012) in which the following four skills for success in elementary school and beyond were identified: Self-control, Mastery Orientation, Persistence, and Academic Self-Efficacy. Similarly, Strive Together in Collaboration with Philliber Research Associates identified five key skill areas through a systematic literature review: Academic Self-Efficacy, Growth Mindset or Mastery Orientation, Grit or Perseverance, Emotional Competence, and Self Regulated Learning and Study Skills (Strive Together, 2013a). However, among all of these approaches, it has been SEL that has attracted the most attention in research, practice, and policy, and hence is the focus in the current study (Durlak, Domitrovich, Weissberg, & Gullotta, 2015).

Policy Regarding SEL

The large body of research supporting SEL as a form of both prevention and promotion has prompted educators and policy makers to examine SEL and related programming for potential use at local, state, and national levels. Both the naturally unfolding and deliberate program-based practices that impact students' social and emotional development already occurring in schools, districts, and states are actively being adapted to better align with evidence-based strategies both in the U.S. and abroad.

Perhaps the most salient example at the state level is the adaptation of learning standards to better include and address SEL. These changes, thus far, have occurred in two ways for grades K-12. Some states have elected to integrate SEL standards into other sets of standards, such as Common Core State Standards, while others have developed comprehensive free-standing, developmentally driven SEL standards (Dusenbury, Weissberg, Goren, & Domitrovich, 2014). Illinois, Kansas, and Pennsylvania have each developed freestanding learning standards that target and identify the development of specific social-emotional skills across the developmental trajectory from kindergarten to high school (Dusenbury et al., 2014). At the preschool level, a recent scan of state standards performed by CASEL (Dusenbury et al., 2014) revealed that 49 states have freestanding standards for social-emotional development, and that many support SEL with accompanying guidelines and resources. Absent of state standards, some school districts have independently developed learning standards related to SEL, such as Anchorage School District in Alaska (<http://alaskaice.org/school-climate/anchorage/>), which among other districts has served as a pioneer in SEL implementation.

At the federal level, there is growing support for SEL through proposed amendments and legislation supported by representatives in congress, such as Susan Davis (CA), Ryan Reps (CA), Dave Loseback (IA), Matt Catwright (PA), John Yarmouth (KY), and Aaron Schock (IL). Two bills were introduced to Congress in 2015. The “Supporting Social and Emotional Learning Act” includes amendments to the Education Sciences Reform Act of 2002 requiring the National Center for Education Research and The Commissioner for Education Research to support and carry out research related to the impact of social emotional skills. In addition, it requires training, professional development, and technical assistance for the use of evidence-based teaching and assessment methods related to SEL to be provided by comprehensive centers of excellence. The

second bill is the “Academic, Social, and Emotional Learning Act of 2015,” which includes a definition of SEL and related programming, identifies core competency areas, and amends the Elementary and Secondary Education Act to allocate funding for educator training and professional development to include SEL programming (CASEL, 2015a).

The Need for SEL Assessment

As systematic research provides a foundation for educators, researchers, and policy makers to advance the use of SEL at local, state, and national levels, the need for accurate and feasible methods of SEL assessment increases (Haggerty, Elgin, & Woolley, 2011). Many authors discuss how the development of such measures is not occurring fast enough to keep pace with the implementation of SEL and related programming, and how essential these measures are for effective SEL implementation. The necessity of measuring student progress toward rapidly developing statewide social-emotional learning standards is commonly noted among these same authors (Kendziora, Weissberg, Ji, & Dusenbury, 2011; Zinsser, Weissberg, & Dusenbury, 2013). Dusenbury (2014) identified that state standards are more likely to successfully impact educational practice when tools, such as assessment, are utilized to support implementation. States that have adopted SEL standards, such as Indiana, address the need for ongoing assessment built on systematic observation, and view it as pivotal in gathering pertinent information that will drive program success and sustainability (Indiana Department of Education, 2012). State-funded quality improvement initiatives (e.g., Heat Start, early special education and child care services) will also likely be influenced by SEL and rely heavily on assessment results to maintain funding (Kendziora et al., 2011). Further, Conley (2014) is among researchers who have called for measures of similar competencies to be used in high-stakes decision-making, such as the college admission process.

Assessment data have an assortment of uses and implications. For example, needs assessments have a tremendous impact in promoting effective program implementation by providing an accurate view of SEL skill development among students in a school, district, or state (Haggerty, Elgin, & Woolley, 2011). As programs are implemented, assessment assists in determining how effectively SEL skills and competencies are being fostered by comparing baseline scores to various time points across subgroups. This can serve to highlight gaps in programming or instruction, as well as identify educators and students in need of additional support. Using assessment to monitor student progress over time at individual and group levels can also add developmentally driven knowledge regarding typical skill progression among various populations or in a particular setting. Additionally, consider that educators equipped with more comprehensive knowledge of students' SEL competencies and the respective course of development in each skill grouping can tailor interventions to meet areas of particular need more efficiently (Bernard, Elias, Bell, Ferrito, & Langione, 2015). Further, by systematically assessing SEL, best practices for supporting SEL skill development can be more readily identified. These best practices are wide in breadth, and can range from individual and group-based intervention, school-wide programming, and national policy.

Current State of SEL Assessment

Multiple large-scale reviews have targeted the identification of which behavioral, social, emotional, diagnostic, and functional measures are most appropriate and effective in assessing SEL (e.g., Denham, Ji, Hamre, 2010; Haggerty, Elgin, & Woolley, 2011). These reviews will likely continue to emerge given the pressing need; however, current trends in the literature highlight numerous complexities in assessing SEL, advantages and disadvantages of different methods of assessment, and recommendations based on the most recent evidence.

The Multifaceted Nature of SEL Skills. Investigations into specific aspects of SEL and related areas have yielded a diverse array of terminology (e.g., “social and emotional intelligence,” “emotional literacy”) and corresponding definitions. The lack of consistent terminology can create barriers to collaboration among researchers and across disciplines as terminology is often utilized or assigned to establish or differentiate areas of investigation (Watson & Emory, 2010). It can also negatively impact construct validity, as underlying constructs and properties of various aspects of SEL have been developed somewhat independent of one another (Watson & Emory, 2010). While the conceptual basis of the constructs and level of construct and content validity is more essential than the precise labels, the investigations carried out by the various disciplines involved (e.g., neuroscience, child development, intelligence theory, organizational learning) have created views of SEL ranging from fixed innate characteristics to flexible and context bound skills (Watson & Emory, 2010).

Wigelsworth, Humphrey, Kalmbouka, & Lendrum (2010) note that attempts at finding commonalities across conceptualizations of SEL initially yielded expansive ranges, which hinder the development of precise scientific construct definitions. Some consensus was reached on overarching trends, such as the importance of competencies (e.g., knowledge attitudes, and behaviors), environments and determinants of SEL development, and skills that were employable and transferable as these were identified across researchers (Watson & Emory, 2010). However, semantic and conceptual differences remained. The work of Denham (2005a & 2005b) reviewing existing measures and developing the initial conceptualization of the five core skill areas currently promoted by CASEL (i.e., the CASEL Five) is highlighted as crucial by many authors (e.g., Wigelsworth et al., 2010) in that it provided a clear framework of SEL skills, which has

been utilized by many researchers to solidify the essential nature of these specific SEL competencies.

Despite the tremendous growth of program efficacy literature supporting the five key skill areas promoted by CASEL, assessment research struggles to keep pace in identifying and developing tools that appropriately assess these five skill areas. One source of difficulty is that measures utilized in some SEL outcome research were developed before or separate from conceptualizations of SEL. These include measures intended to recognize problematic behaviors (e.g., aggression), social skills (e.g., eye contact, sharing), emotional difficulties (e.g., mood lability), functional skills (e.g., activities of daily living, hygiene), and psychiatric diagnoses (e.g., anxiety). These measures were used because some constructs assessed were viewed as related to SEL or targeted outcomes believed to be influenced by SEL (e.g., decrease in conduct behaviors). Further, difficulties also ensue in the use of measures from fields related to SEL (e.g., emotional intelligence, personality trait development) because these measures were based on differing terminologies, definitions and conceptualizations of what is now enveloped by SEL.

Wigelsworth et al. (2010) note that differences in terminology and conceptualizations across measures intended for SEL and related fields may not be as significant when viewed in comparison to similarities thus making many terms interchangeable. They provide several examples of how these similarities manifest on current measures, including measures that conceptualize SEL differently, such as a trait (i.e., stable personality characteristic) perspective as opposed to a skill perspective (i.e., behaviors viewed as more susceptible to change). More systematically, similar approaches have been undertaken in the publication of compendia by organizations such as CASEL (Denham et al., 2010) and The Raikes Foundation (Haggerty et al., 2011), which map domains assessed by existing measures, including those designed entirely

separately from SEL and related fields, onto the five core skill areas held by CASEL. These reviews contribute tremendously to bridging gaps and mitigating differentiations among similar concepts previously divided due to terminology; however, they also show that there are few measures in existence that assess all five of the core skill areas.

Psychometric Properties. There is noted variability in the reliability of scores and validity of inferences drawn from those scores on SEL measures and assessment in related areas (e.g., Humphrey et al., 2011). In Durlak et al.'s (2011) systematic review of 213 school-based universal SEL interventions, authors defined a measure as reliable if kappa or alpha statistics $\geq .60$, reliability calculated by product moment correlations was $\geq .70$, and level of percentage agreement by raters was $\geq .80$. A measure was considered valid if data confirming the construct, concurrent, or predictive validity was cited. With these standards, 24% of studies reviewed were found to use measures that did not yield reliable scores and 49% of studies reviewed were found to use measures from which valid conclusions could not be made. The large percentages of measures lacking appropriate psychometric properties found through this review is concerning. In a research and development plan by the RAND Foundation entitled "Measuring Hard-to-Measure Student Competencies," potential harms of using measures that lack high technical quality are raised along with how the paucity of sound assessment instruments may negatively impact the willingness of educators, parents, and other decision makers to support the use of such measures in educational settings (Stecher & Hamilton, 2014).

Identifying measures with strong psychometric qualities has been part of the goal of several large scale reviews of measures. CASEL developed a compendium of more than 50 measures for preschool through elementary school students (Denham et al., 2010). For every measure, both reliability and validity statistics were provided, basic criteria were met, and

strengths and weaknesses discussed. Similarly, The Raikes Foundation reviewed 73 measures and created a final compendium of 10 recommended measures for middle school students (Haggerty, Elgin, & Woolley, 2011). In order for a measure to be recommended, reliability and validity criteria had to be satisfied along with criteria related to suitability for program evaluation, ready availability of the measure and relevant information to schools, and the measure needed to not have been designed to assess specific programs. Internal consistency and test-retest reliability needed to be .70 or above, and strong evidence for criterion-related validity, and convergent and/or discriminant validity needed to have been established (Haggerty, Elgin, & Woolley, 2011).

The numerous measures identified through these reviews indicate that measures with relatively strong psychometric properties exist; however, the nature of SEL as a diverse field with both differing and evolving conceptualizations highlights the necessity for similar reviews to continue to occur. This need is underscored in that social-emotional needs of children can be misunderstood if culture and perspective are not appropriately considered (Hudley, 2001). This is likely because culture provides a context through which children develop a sense of identity and framework that assists them in understanding the world (Denham & Weissberg, 2004). This context is viewed as constantly evolving through both direct (e.g., family, school, friend), and indirect (e.g., media, social services) influences (Barblett & Maloney, 2010) driving the notion that work across cultures and settings needs to be rigorous.

Scope, Specificity, Purpose, and Setting. Practical complexities emerge in the need to match measures appropriately with particular purposes and settings. McKown (2015) noted that there are many purposes to the assessment of any construct and made distinctions between assessing individuals or the environment, processes or outcomes, teacher practices or student

behaviors, and programs. Denham (2015) similarly discussed differences between screening formative, interim, and summative assessment approaches. Screening allows cutoffs to identify students in need of additional assistance. Formative serves as an ongoing source of feedback on student progress that is integrated into teaching and learning practices. Interim provides data at the classroom, school, or district level one or more times in an academic year, and summative serves as an assessment of learning at given time intervals (e.g., marking periods) that evaluates student performance against content standards (Denham, 2015).

Clearly each of these forms of assessments has implications in terms of feasibility (e.g., cost, completion time, existence of SEL content standards for that school etc.). For instance, consideration of issues related to resources (e.g., cost of measures, scoring procedures, who serves as the respondent, methods of data collection) (Wigelsworth, et al., 2010) differ significantly if a measure is to be used universally (e.g., throughout an entire school) as a screening tool to identify students at-risk, as an ongoing formative assessment, or as a summative assessment conducted twice a year. The emphasis on immediate feedback in formative assessment, for example, may necessitate that measures be quick for educators to administer and interpret, whereas summative assessments may afford additional time and resources for administration and scoring in the interest of comprehensive assessment. Similar considerations are relevant if a measure is intended for targeted use (e.g., identifying students who require additional support; monitoring progress of an intervention) because although some of these issues may be less prominent with this approach (e.g., cost), others remain essential in both targeted and universal approaches (Wigelsworth et al., 2010). As discussed above (see Psychometric Properties), regardless of the number of students assessed, the cultural composition

of the population assessed will remain a salient factor as many tools were developed or offer norms based on specific subpopulations.

Additionally, the scope and specificity of measures and the domains assessed are essential and need to be matched to both purpose and setting (Wigelsworth et al., 2010). Brief and time efficient measures (e.g., as few as eight items as seen on the Devereux Student Strengths Assessment-Mini) (LeBuffe, Shapiro, & Naglieri, 2012) tend to lack specificity, be one-dimensional, and be less sensitive to change over time (Wigelsworth et al., 2010).

Conversely, multidimensional measures capable of providing more detailed information and more effectively tracking change over time tend to be lengthier (e.g., upwards of 140 items as seen on the Social-Skills Improvement System) (Haggerty, Elgin, & Woolley, 2011; Wigelsworth et al., 2010).

Types of Measures

There are three dominant types of measures distinguished primarily by the source and process of gathering information: performance-based assessments, student self-report, and observer-report. Advantages and disadvantages of each are discussed in the sections below, as well as recommendations for selecting and developing measures. Knowing the broader categories in which each of these and other measures fall can be pivotal in understanding the general theory supporting the measure and in examining each approach for potential use. In Lipton and Nowicki's (2009) discussion of SEL assessment theory, the authors distinguish between measurement of the execution of SEL and related behaviors, and comprehension of social-emotional information including encoding, interpreting, and reasoning abilities. Self-report and observer-report approaches are better suited to measure execution, while performance-based assessments are more appropriate measures of comprehension (McKown, 2015). It is

important to note that direct observation is another execution-focused method of assessment commonly found in the literature. Some authors highlight the importance of capturing the context in which SEL and related behaviors emerge rather than reporting the general frequency as is found in self and observer report measures. The significant time and resources necessary to systematically observe student behavior across contexts and establish reliability, however, create tremendous barriers to recognizing this as a feasible and scalable approach (McKown, 2015).

Performance-Based Assessments. Similar to how cognitive and achievement abilities are assessed in schools, some researchers advocate for the use of performance-based assessments of SEL and related skills; particularly in the interest of measuring comprehension as discussed above (McKown, 2015). Included in these assessments are tasks such as identifying emotions displayed in images of faces and identifying the emotions evoked in the examinee by a protagonist character in a short story. The foundation of these measures is that they are direct in tapping underlying constructs through such tasks (Humphrey et al., 2007), and may potentially have robust validity as a result (Kendziora et al., 2011).

Performance-based assessments, however, have significant practical disadvantages. The need for “expert” personnel to administer, score, and interpret measures, as well as relative cost compared to report-based measures create strains on resources (e.g., time, financial implications). These serve as large barriers to widespread adoption and use. Additionally, claims of the potential for increased validity when compared to other measures are disputed by some authors who identify that “correct” answers on these tasks may be difficult to identify and rely on expert and/or consensus opinion which are susceptible to cultural bias (Wigelsworth et al., 2010). Further difficulties arise in the need to use multiple measures as many existing performance-based measures only address one dimension of SEL (McKown, 2015). McKown (2015) asserts

that research continues to be conducted in developing feasible and scalable methods of performance-based assessments that entail core dimensions of SEL. In the interim the authors suggest the use of such assessments be reserved for indicated cases in conjunction with other measures (McKown, 2015).

Student Self-Report. Many researchers and teachers identify that including the student's perspective in the assessment process is important (Barblett & Maloney, 2010; Kendziora et al., 2011). Advantages identified in the literature include that through introspection each student has access to the most detailed information about him/herself (Wigelsworth et al., 2010); particularly for competencies that are not as readily observable such as internal processes or symptoms (Elliot, Frey, & Davies, 2015). This may be particularly relevant for aspects of SEL that are not as readily observable (e.g., Self-Awareness). The positive practical implications of student self-report methods are also acknowledged with particular emphasis on the feasibility and scalability of SEL measure use in schools (Wigelsworth et al., 2010). The use of self-report measures as screening tools has also been supported (Elliot, Frey, & Davies, 2015). There are clear benefits related to time and resources in utilizing self-report measures; particularly as the population of students being assessed increases.

Disadvantages of self-report methods most notably involve developmental differences across ages that may negatively impact validity and reliability; particularly for students in pre-K through fifth grade (Kendziora, et al., 2011). For example, self-awareness and self-perception follow a developmental trajectory making older adolescents better able to provide accurate responses than younger children (Denham, 2005b). Also, younger children may be more likely to be biased toward basing responses on recent events rather than accounting for experiences over time (Wigelsworth et al., 2010). For instance, an item such as "I get along well with others" may

be more likely to receive a lower rating if a younger child recently had an argument with a friend; even if the two typically get along well (Wigelsworth et al., 2010). Additionally found in the literature is the notion that children in particular may be more likely to give socially desirable responses (Wigelsworth et al., 2010); however, given that items on SEL self-report measures tend to be face-valid, it may be that there is a certain level of risk for responses to be biased across age ranges. Cognitive, language, and reading abilities also play a critical role in influencing the collection of valid data (Elliot, Frey, & Davies, 2015). These factors may limit the use of such measures with certain students (e.g., younger students, those with specific learning disabilities), or require certain students receive support in completing these measures which raises additional issues (e.g., how responses can be influenced by whoever assists).

Observer-Report. Parents, peers, and teachers have all been identified as potential sources of information. Advantages of parent-report measures most notably entail the opportunity for observations of the student at home to provide a different perspective (Wigelsworth et al., 2010). Disadvantages include that parents have a more restricted frame of reference in comparing a child to other students of that age, and that difficulties in recruiting and retaining parents found in the literature can serve as barriers to intervention and evaluation efforts (Wigelsworth, et al., 2010). For example, Wigelsworth et al. (2010) cited a study conducted by Humphrey and colleagues in 2008 where parents were found to have a 50% initial response rate, with additional drop outs at each sequential data collection time point.

Peers have emerged as potential raters dominantly through research related to social skills identifying a potentially advantageous evidence base from which to build (Wigelsworth et al., 2010). However, issues have emerged in that evidence base related to the influence of personal

attributes such as attractiveness and academic success. Also, interactive factors such as similarities in gender and race may also skew results (Wigelsworth et al., 2010).

Some advantages of teacher-report measures flow naturally from disadvantages found among other reporters. For example, unlike parents, teachers generally have consistent exposure to many students around the same age enabling comparisons and similarities to be identified more readily (Kendziora et al., 2011; Wigelsworth et al., 2010). Also, knowledge of student behavior in an important environment (i.e., school) is valuable (Kendziora, Weissberg, Ji, & Dusenbury, 2011). Additionally, numerous measures based on teacher report have been explored and established reliability and validity along with methods of streamlining administration and interpretation (e.g., benchmarks and standards provided with measures, availability of electronic administration and scoring) (Kendziora et al., 2011).

Disadvantages are also present in teacher-reports. Teacher bias and assessment subjectivity have emerged related to each teacher's cultural lens (Barblett & Maloney, 2010), and the extent to which their own social-emotional needs influence the attitudes and behaviors they recognize and transmit to students (Weare & Gray, 2003). Additionally, some studies suggest that behaviors displayed less frequently may be more likely to yield biased and negatively slanted behavior ratings by teachers (Christ, Riley-Tillman, Chafouleas, and Jaffrey, 2011). The extent and nature teacher training and experience (e.g., completing measures, observing student behavior systematically, working at that grade level) can also influence ratings (Wigelsworth et al., 2010). For example, Denham (2005b) found more experienced teachers tended to provide higher ratings of SEL related behaviors. Also, teacher workload and philosophy toward formal and informal assessment methods may influence data as measures can be time-consuming (Edmunds & Stewart-Brown, 2003).

Multiple-Informants. Some authors advocate that best practice involves multiple informants rating the same child's behavior to provide a more complete view of behavior across situations and settings (Elliot, Frey, & Davies, 2015). The ability to triangulate data across several informants is recognized as a method to promote accuracy (Wigelsworth et al., 2010), and identify behaviors that manifest across a variety of situations as opposed to those that are more situation specific (Elliot, Frey, & Davies, 2015). Although there are clear advantages, researchers have repeatedly found that multiple informants often demonstrate moderate agreement at best (Elliot, Frey, & Davies, 2015; Gresham, Elliot, Cook, Vance, & Kettler, 2010; Renk & Phares, 2004) and enhancing agreement often involves additional measurement strategies (e.g., systematic behavioral observation). Issues related to resources such as cost of multiple measures per student, time allocated to distributing, collecting, and scoring measures, recruiting and retaining informants (e.g., parents), and so on indicate that there are disadvantages from a feasibility standpoint.

Another potential method for integrating information from multiple informants is through formative assessment. Teachers could readily integrate qualitative or quantitative (e.g. ratings based on a daily rubric detailing stages of skill development) information gathered through probing discussions or unobtrusive assessments (e.g., observing students in a disagreement) with student self-assessments (e.g., self-ratings on a daily rubric). The ongoing collection of data from both students and teachers can provide some of the advantages discussed above, as well as potentially enable a platform for students to track their progress based on data from two informants (Marzano, 2015). However, these processes require additional time and resources, as well as expertise for the teacher to be able to integrate and utilize this information in a meaningful manner (e.g., to instruct teaching practices or programming; to provide ongoing

methods to track student progress; to provide meaningful feedback or develop systems for students to track progress) (Marzano, 2015).

Recommendations for Selecting SEL Measures. Many authors acknowledge difficulties in appropriately selecting measures, particularly in SEL and related fields given the complexities outlined above, and some offered guidance to researchers, educators, and policymakers on key considerations when selecting from existing measures. Two of the most salient guiding principles were put forth by Snow and Van Hemel (2008) who served as editors for the National Research Council after the U.S. Congress requested a study of developmental outcomes and appropriate assessments of young children. The two guidelines overlap with recommendations from an array of authors (e.g., Crowe, Beauchamp, Catroppa, & Anderson, 2011) and are as follows: (1). Purposefulness, which holds that measures be selected based on how it will be used (e.g., to evaluate programs, assess a child’s functional capacities); (2). “Systematicity,” which holds that assessments are only given in a context of care and educational supports that can constructively use the data to promote children’s optimal development (Snow & Van Hemel, 2008).

Kenziora and colleagues (2011) also produced five principles for SEL assessment. They focused on how assessment was recognized as a potentially demanding process (e.g., time, energy) and encouraged awareness of how staff may become overburdened or anxiety among students may rise. They stressed that the primary goals are to better understand children’s strengths and areas of needs, improve students’ SEL skill and competencies, and to help students reach successful academic and life outcomes. These considerations, taken together, suggest the importance of a feasible and scalable approach to SEL assessment that can be integrated into ongoing mechanisms for feedback and student skill development.

Measuring SEL Competencies on ‘The Other Side of the Report Card’. Ultimately, the success of any initiative or program is based on the method of implementation. The work of several authors (e.g., Denham, 2015; Kendziora et al., 2011) highlights how barriers related to resources can negatively impact the production of meaningful information able to drive program and teacher efficacy, and student success even among those measures which meet many of the above recommendations. Consider, for example, a measure with strong psychometric properties and appropriateness for the population and purpose. If factors such as the time it takes to complete the measure, cost, scoring and interpreting procedures, ease of use, and mechanisms of delivering feedback to teachers, parents, and students are not well-addressed, then the quality and utility of the data collected will be negatively impacted. Although no particular assessment system has been identified that adequately addresses barriers to feasibility and scalability, there is one practice that emerged based on the original intuition from educators that behaviors matter: report card comments.

Generally, the system for assigning letter grades is consistent across schools. However, report card comments are far from systematic and can vary tremendously across schools, at times even within the same district. Comments often serve as the only uniform feedback given by school personnel to students and parents regarding behavior in school. Interestingly, these comments often already include aspects of SEL skills, albeit in a manner that is highly variable and lacking an evidence base. Considering the pervasiveness of report card comments, very little research has been devoted to their structure. In fact, through the work at the Social-Emotional Learning (SEL) Lab at Rutgers University directed by Dr. Maurice Elias, it has been found that many districts are unclear on how report comments came to be selected and implemented. Many district administrators and educators held that comments emerged historically and were unaware

of any systematic approach, evidence base, or theoretical rationale to the development or adaptation of comments (M. Elias, personal communication, June 22, 2015).

Stephen J. Friedman led two research teams in Wisconsin in some of the only recently published research on report card comments prior to the work of the SEL Lab at Rutgers University (Friedman & Frisbee, 1995; Friedman, Valde, & Obermeyer., 1998). The studies examined characteristics of report cards (Freidman & Frisbee, 1995) and teachers' use of computerized report card comments (Friedman et al., 1998). Approximately 39 kindergarten, 59 elementary, 48 middle school and 70 high school report cards were analyzed. At no grade level were statements of philosophy or purpose included with any great frequency, and as grade level increased, statements related to philosophy and purpose decreased in frequency (Friedman & Frisbee, 1995). Formats ranged from unstructured space for teachers to write comments to computerized drop down menus of often as many as 80 different possible comments from which teachers were allowed to select two for each student (Friedman & Frisbee, 1995).

Friedman and colleagues' second study (1998) focused on computerized report card comments through examination of 475 student report card comments in a small town in Wisconsin. Teachers were able to select two comments from a menu of 82 possible comments. Only 52% of students received two comments and teachers ($n = 37$) on average used a total of 17 different comments for all students, a mere 20.7% of their options, indicating that 82 options may be impractical. Teachers were approximately three times more likely to provide positive behavioral or academic comments overall despite almost three times more negative than positive comment options. Also, teachers tended to use negative behavioral or academic comments to explain low grades rather than positive comments to explain high grades. While many teachers (72%) felt that offering two comments per student was "about right", the majority of parents

(58%) disagreed saying this was not enough. Overall, the majority of teachers and parents found the comment system to be “somewhat helpful” and agreed it could benefit from revision contingent upon clarity of the intended purpose of comments (e.g., rewarding achievement, informing parents, ranking students, motivating students).

The “Other Side of the Report Card” (Elias, Wang, Weissberg, Zins, & Wallberg, 2002) refers to the behaviorally-driven comments included on nearly every report card. At the Social-Emotional Learning (SEL) Lab at Rutgers University, the paucity of research on styles of feedback and behaviors included in report card comments presented an opportunity to utilize an existing system to which teachers already allocate time, and for which districts already provide funding. Moreover, mechanisms for sharing information with students and teachers and parents are already in place, and student progress can readily be tracked to report on meaningful SEL skills. Three studies emerged from the SEL Lab at Rutgers University to investigate the relationship between report card comments and academic achievement at the high school, middle school, and elementary school levels.

Kemp et al. (2014) examined the relationship of race/ethnicity and gender with behavioral comments and letter grades from three middle schools in a large suburban district in New Jersey. These schools included grades six, seven, and eight and analyses focused on Hispanic, Black, and White students as other ethnicity groups were not represented by enough students for comprehensive statistical analysis. Within these ethnicity groups 200 students from each of the three grade levels were randomly selected making a total sample size of 600 students with 54.7% being female, 64.7% being White, 19.7% being Black, and 15.7% being Hispanic. Teachers selected comments from a list of 16 with a limit of three comments per class per marking period. Comments were identified as positive (e.g., shows consistent effort) and

negative (e.g., does not follow directions). Significant differences by race/ethnicity and gender were found for comments and letter grades. Black, Hispanic, and male students received more negative comments, less positive comments, and lower letter grades than White and female students. The reason for differences could not be derived from this investigation, however, authors cautioned school personnel about potential rater bias and the need to address behavioral discrepancies among subgroups in schools.

Moceri et al. (2014), using a sample of over 1,000 students from an ethnically diverse, large high school, investigated the relationship between existing comment sections, SEL skills, attendance, academic grades, and standardized test scores to determine if comments related to SEL were indicative of current and future academic success. Report cards featured 25 potential comments from which teachers selected two comments per class per marking period. Comments only partially assessed elements of the CASEL 5 with 10 of the 25 comments being qualitatively categorized by the research team as related to SEL core competencies. Both SEL and Non-SEL related comments were divided into positive and negative groups.

Moceri et al. (2014) focused on demographics for each student in initial analyses. Approximately half of the sample was female, 58% Black, and almost 20% qualified for free and reduced lunch. Black students, male students, and students qualifying for free or reduced lunch had more negative comments and lower standardized test scores than peers who were White, female, and did not qualify for free or reduced lunch peers. Additionally, Black students had fewer positive comments and lower grades than White students.

After controlling for demographics, comments were found to have a small effect size on attendance, and a large effect size on current letter grades in both “subjective” (i.e., language arts) and “objective” (i.e., mathematics) subjects. Negative comments were found to have two to

three times greater of an effect on letter grades than positive ones. Additionally, after accounting for report card comments, the effect of demographics on letter grades decreased dramatically. Specifically with regard to Language Arts, the effect of gender reduced by roughly 60% and the effect of ethnicity decreased by approximately 40%. For math, the effect of gender decreased by approximately 85% and the effect of ethnicity reduced by approximately 50%. These findings support the link between report card behavior ratings, including those linked to SEL skills, and academic report card grades. Standardized test scores were only available for students in eleventh grade and analyses revealed that comment categories were able to explain 11% of the variation in Language Arts and 16% of Math standardized test scores. Findings that comments were unable to explain standardized test scores or to predict future letter grades after accounting for prior letter grades were explained by the strong relationship between comments and letter grades.

Moceri and Elias (2014) aimed to replicate the work of Moceri et al. (2014) with 186 4th and 5th grade students where approximately half the students were female, nearly a third qualified for free or reduced lunch, and half the students were Black, Hispanic, Asian, or multiracial. This report card format differed in that teachers assigned comments from a list of 24 for each academic subject for each marking period with no limit to how many comments could be assigned. Slightly more than half ($n = 13$) of the 24 comments were related to two of the CASEL 5 core potencies of Self-Management or Relationship Skills. Comments were broken down on the report card into positive and negative. Three common groupings for frequency per class per marking period and type of comment assignment labeled as Low, Medium, and High. Students in the Low cluster received a small number of negative and positive comments (i.e., on average less than two positive and/or negative comments). Those in the Medium cluster were

assigned a moderate number of positive (i.e., on average six) and low amount of negative comments (i.e., on average one). In the High cluster, students were assigned many positive comments (i.e., on average eight), and rarely a negative comment (i.e., on average less than one).

Strong evidence linked positive comments to better letter grades in “subjective” (i.e., language) and “objective” (i.e., math) subjects, and some evidence linked positive comments to better performance on standardized tests. On average, students who received medium or high amount of positive comments received a third of a letter grade higher than those who received a low amount. Additionally, students who received more positive comments were more likely to achieve advanced proficiency on standardized tests.

Moceri et al. (2014) noted several limitations, many of which are relevant to each of the above studies, and provided recommendations for future research. Restrictions on how many comments a teacher could assign per student per class per marking period limited the comprehensiveness analyses, particularly because comment variables were dichotomous (i.e., present or absent). Comments’ limited representation of the CASEL 5 core competencies of SEL was identified as another significant hindrance in examining the potential predictive validity of SEL related comments. Additional limitations included that the same raters simultaneously assigned letter grades and report card comments with no psychometrically acceptable behavior scales for comparison or ability to conduct teacher or class level analyses due to restrictions in the data. Recommendations for future research included the use of psychometrically sound behavioral measures to check the validity of report card comments and look at teacher level effects, examination of schools with various grade configurations, and further investigation of demographic variables including SES and ethnicity. Also, additional variance in response

choices was encouraged by the author, potentially through the use of a short set of Likert rating items so as to increase the probability of predictive validity analyses.

The Current Study

The current study builds on the prior studies by Kemp et al. (2014), Mocerri et al. (2014), and Mocerri and Elias (2014). It aims to further investigate report card comments relationship to academic achievement through an SEL lens consistent with the CASEL 5. The relationship of typical report card comments representative of certain SEL skills and academic achievement will be investigated. This study follows the recommendations of Mocerri et al. (2014) in that Likert ratings of specific behaviors based on early social-emotional skills theory are present in the report card comment section in addition to typical comments. This is significant as comments and rating items are expected to more comprehensively address the CASEL 5 in comparison to comments examined by Mocerri et al. (2014). Also, Likert rating items allow for more variability in responses and may bolster the accuracy of analyses. The work of Kemp et al. (2014), Mocerri et al. (2014) and Mocerri and Elias (2014) is promising regarding the relationship between SEL related comments and academic achievement given the dichotomous nature of typical comments and limited representation of the CASEL 5. This study aims to build upon that foundation and may have a more robust representation of SEL and greater sensitivity in measuring SEL related competencies through use of the Likert rating items.

This study provides the first investigation of a behavior rating style of report card comments specifically developed to address social-emotional competencies. Additionally, this study provides the first, to the knowledge of this writer, side by side comparison of typical report card comments with rating items geared toward SEL implemented with the same sample population on the same report card. As educational policy and State Standards continue to

mandate instruction of SEL and related skills, assessment will undoubtedly follow as is discussed above. The examination of a brief Likert rating system in practice on a report card, rooted in SEL, and developed by the district, highlights the potential for similarly adapted comment systems to offer a usable, feasible, acceptable, and scalable approach to assessing SEL in schools. These advantages are critical given the array of barriers to SEL assessment discussed above and that typical SEL research involves an outside measure being distributed, collected and interpreted which has numerous implications, particularly with regard to feasibility for practice and resources (e.g., cost, time of expert personnel for scoring and interpreting). This study focuses on how well comments and SEL-based Likert rating items align with current SEL theory (i.e., CASEL 5), and the concurrent validity of both comments and ratings for academic achievement as measured by academic grades and standardized test scores.

RQ1: Which SEL skill categories are represented by the 24 typical report card comments? Which SEL skill categories are represented by the seven “Personal and Social Development” rating items? What conceptual categories can be created based on trends in the assignment of the 24 typical report card comments? What conceptual categories can be created based on trends in ratings on “Personal and Social Development” items?

RQ2: What student behaviors observed (i.e., typical comments selected from the list of 24) and/or rated by teachers (i.e., seven “Personal and Social Development” items), related to SEL skill categories, are associated with academic achievement? Do any SEL skill categories represented by teacher observations and/or ratings of these behaviors have a stronger relationship with high academic achievement than other SEL skill categories? Do any student behaviors distinct from SEL skill categories observed and/or rated by teachers have a stronger relationship than other behaviors with academic achievement?

H2a: Students who received the most comments from the list of 24 representing the presence of behaviors related to SEL skills and highest ratings on the seven items addressing SEL skills are predicted to demonstrate the highest academic achievement. This is consistent with findings from the most recent meta-analysis of SEL programming that students involved in programming targeting the development of SEL skills showed significantly higher academic achievement (Durlak et al., 2011).

H2b: Students who received the most typical comments from the list of 24 and highest ratings on items representing the presence of behaviors related to the SEL skill category of Social-Awareness and Relationship Skills are predicted to demonstrate the highest academic achievement. In each of the core academic areas, ratings on sub-domains pertaining to each subject were provided on the report cards used in this study (see Appendix B). Aside from specific academic skills (e.g., “uses proper punctuation”), these sub-domains involve participation, group work, and ability to communicate subject area material. These sub-domains drive this prediction as these behaviors can be conceptualized as representative of Social-Awareness and Relationship Skills given that they require communication skills and are reinforced by social consequences (e.g., parent and teacher reactions).

Chapter II

Method

Setting and Participants

One-hundred-thirteen fourth grade students from a suburban elementary school in New Jersey with a total enrollment of approximately 400 students participated in the current study. Demographic data from the NJASK, administered during the 2006-2007 academic year, provided the most representative retrospective breakdown of race and ethnicity for the fourth grade population. Approximately 53% of students were female with an estimated 49% of all students identifying as White, 21% Black, 18% Asian/Pacific Islander, and 12% Hispanic. During the initial data collection process, an undergraduate student in the Social-Emotional Learning Lab at Rutgers University assigned each student an identification number and the data were de-identified when provided to this author for the current study. Table 2 depicts demographic information for the sample population.

The school used for this dissertation is appropriate for this study because it: (a) features both a list of typical comments and specific items that provide ratings of SEL, (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 specific aspects related to a unique program, (c) has no formal or consultative relationship with CASEL.

Inclusion criteria for this study's analyses were all students enrolled in the 4th grade during the 2006 to 2007 academic year who received English and Math scores on the NJASK in the spring of 2007, as well as subject area grades for each of the four marking periods. Exclusion criteria were students of other grade levels (students enrolled in second, third, and fifth grade compose approximately 75% of the school's total enrollment). At the time of the study, NJASK

was the standardized academic test in mandatory use in schools in New Jersey. Students enrolled in second grade did not take the NJASK and were thus excluded. Additionally, students in both second and third grade did not receive total scores or grades in each subject area, but rather ratings on approximately six sub-domains within each subject. Without a total subject area grade, secondary analyses were not possible for students in second and third grade, and those students were subsequently excluded. NJASK data were not available for students enrolled in fifth grade during the 2006 to 2007 academic year, and those students were also excluded as a result. Reading, writing, math, and science class grades were chosen to contribute to cumulative GPA, as they are the core academic subjects taught in this elementary school in fourth grade.

Measures

Report Card Variables. Academic letter grades, standard comments from a list of 24, and ratings of behaviors related to “Personal and Social Development” are all provided on the report card. Letter grades are provided separately for academic subject areas of reading, writing, science, 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- = .67, F = 0.

Each student can receive as many of the 24 standard comment options (see Table 3) as a teacher chooses to designate. Students receive comments in each of the four core subject areas for each of the four marking periods. These standard comments serve to provide feedback about a wide range of behaviors and are broken down on the report card into two primary categories: positive comments, referring to the presence of a behavior, and negative comments, referring to the need for a behavior to be demonstrated more frequently or consistently.

Students also receive ratings (i.e., N/A, check minus, check, check plus) on seven “Personal and Social Development” items for each of the four marking periods (See Table 4)

where N/A = 1, check minus = 2, check = 3, and check plus = 4. No students received a rating of N/A for any “Personal and Social Development Items.” All of these items are representative of behaviors related to the CASEL skill categories. These items were developed through collaboration among school staff members after implementing an evidence-based SEL program for several years (i.e., Social Decision Making/Social Problem Solving Curriculum). Given that each of these items was based on an SEL curriculum (Elias & Bruene, 2005), some involve language specific to that program. Specifically, “Keep Calm” refers to a self-calming strategy involving breathing and self-talk. The item, “B.E.S.T.” is an acronym referring to a student’s body posture, eye contact, saying the right words, and tone of voice; each of which represent a key skill when interacting with others. “Speaker Power” is an object that, when being held by a classmate. Teachers or staff, indicate it is their turn to speak and that others should be respectful. “Listening Position” involves positions consistent with good listening, including looking at a speaker, staying seated with feet on the floor, and facing a speaker.

Non-Report Card Variables. Students’ gender and standardized test scores were provided from students’ school records. The New Jersey Assessment of Skills and Knowledge (NJASK) was at the time of the study New Jersey’s standardized achievement test for elementary school students. Scaled scores range from 100 to 300 where 100-199 means Partially Proficient, 200-249 means Proficient, and where 250-300 means Advanced Proficient. Students take the NJASK between March and May of each academic year; therefore, standardized test score analyses are limited to this grade level.

Effect Sizes. Cohen’s (1992) classifications of effect sizes will be used as guidelines for the strength of correlations. As defined by Cohen (1992), values for effect sizes are as follows: small ($r = .10$ or $r = -.10$), medium ($r = .30$ or $r = -.30$), and large ($r = .50$ or $r = -.50$). Extending

Cohen's work, further ranges as defined by Hopkins (2002) include values for nonexistent correlations ($r = .00$), very large correlations ($r = .70$ or $-.70$), and nearly perfect correlations ($r = .90$ or $r = -.90$). A change in R square value of .02 is small, .13 is medium, and .26 is large for multiple regression (Cohen, Cohen, West, & Aiken, 2002).

Chapter III

Results

Table 4 provides an overview of data analytic techniques utilized in this study.

RQ1 Face and Content Validity

Qualitative Categorization. To answer the first research question, comments from the list of 24 and the seven “Personal and Social Development” rating items were qualitatively categorized by this author. Reliability was assessed by having an expert in the field of SEL familiar with the development of the CASEL skill categories independently qualitatively categorize comments and rating items. This author and the expert in the field remained blind to one another’s categorizations until completed and compared to reach consensus. Qualitative categorization by both this author and the expert in the field involved identifying which behaviors described in the typical report card comments (see Table 6) and “Personal and Social Development”(see Table 7) items were conceptually consistent with specific skill definitions and observable behaviors described in three overarching SEL skill categories as defined by CASEL: (1) Self-Awareness & Management, (2) Social-Awareness and Relationship Skills, (3) Responsible Decision-Making (See Appendix A). Additional conceptual categories were investigated for any comments not qualitatively categorized as related to the CASEL skill categories. A “Non-SEL Behaviors” category was utilized for any comments that could not be more meaningfully conceptually defined.

After independent, blinded categorization, agreement was found to be approximately 90% for comments and 100% for “Personal and Social Development” items. Consensus was reached after discussion and reevaluation of the CASEL 5 and relevant skill examples (see

Appendix A). The results of categorization for typical comments are depicted in Table 5 and Table 6 depicts results for “Personal and Social Development” items.

Typical comments differed from “Personal and Social Development” items in the process of qualitative categorization, in that comments were not directly based on the school’s SEL programming. As a result, conceptual linkages were not as readily defined between behaviors addressed by comments and SEL skills from a face validity standpoint. In categorizing behaviors described by typical comments, SEL skills necessary to perform the behavior were considered in order for categorization rationales to be apparent. As is true for many behaviors exhibited by students, multiple sources of influence can be identified. This author and the expert in the field focused on the potential impact of the presence or absence of underlying SEL skills on each behavior described by comments. For some typical comments less readily related to SEL, categorization was derived based on expectations and feedback commonly provided to students regarding these behaviors from teachers, and, potentially, caregivers, leading to categorization in the Social Awareness and Relationship Skills category. For example, items such as “brings materials to class daily,” “completes homework regularly,” and “completes classwork regularly,” are representative of behaviors for which students often receive feedback communicating the importance of these behaviors. The persistent lack of attending to these behaviors despite ongoing social cues to engage in them could involve a variety of SEL skills under the Social Awareness and Relationship Skill domain, including listening and considering a different point of view, understanding facial, verbal, and situational cues (e.g., tone of seriousness, seeing other students’ completing work), accurately assessing intentions and handling criticism (e.g., understanding feedback is intended to promote student academic performance), and cooperation.

Exploratory Factor Analysis. Exploratory factor analysis (EFA) was conducted in SPSS to determine if an underlying structure was present in the data based on how comments and, separately, ratings co-occurred, as well as to identify organizational loadings of comments and items to emergent factors. Emergent factors were then qualitatively examined based on item/comment content, and the corresponding subject area and marking period for which it was assigned. Prior to conducting the EFA, the data set was assessed for suitability for the analyses using the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO). A KMO greater than .70 indicates that factor analysis is appropriate for the data set (Meyers, Garnst, & Guarino, 2006), and the KMO for each of the current analyses surpassed that criterion. Principle components factoring with a Varimax rotation was utilized for comments, and again for “Personal and Social Development” items.

EFA for the 24 comments was limited by the low frequency with which many comments were assigned. Table 8 provides a depiction of how frequently comments were assigned on average across marking periods and subject areas. As a result, only those comments provided frequently enough could be included in the analysis. On average, 14 of the 24 comments were offered less than one time across all marking periods and academic subjects. Ten of the twenty-four comments were offered four times or more on average across subject areas and marking periods, all of which were positive comments. EFA of the 10 positive comments was conducted using total scores across academic subjects for all four marking periods, for marking period one, and for marking period four to examine if factors remained consistent. EFA of the 10 comments using total scores summed across academic subjects for all four marking periods (KMO = .93) yielded a one factor model that accounted for 87% of the variance, with the eigenvalue for Factor 1 = 8.65. Similar results were found at marking period one (KMO = .91) and marking period four

(KMO = .93) which both yielded a one factor model that accounted for 74% and 81% of the variance respectively. Factor loadings using total scores across academic subjects and marking periods for the 10 positive comments are depicted in Table 9.

EFA for the seven “Personal and Social Development” items was conducted using total scores for each item summed across all four marking periods, for marking period one and for marking period four to examine if factors remained consistent. EFA of the seven Likert items using total scores across four marking periods (KMO = .88) yielded a one factor model (see Table 10) that accounted for 75% of the variance, with the eigenvalue for Factor 1 = 5.24. EFA of the seven items for marking period one (KMO = .89) yielded a one factor model that accounted for 66% of the variance. For marking period four, EFA of the seven items (KMO = .90) also yielded a one factor model that accounted for 78% of the variance.

Data Reduction. EFA results at multiple time points yielded one factor models that accounted for 74-87% of the variance for the 10 positive comments and 66-78% of the variance for the seven “Personal and Social Development” items. To provide further evidence of reliability and inform data reduction strategies, Cronbach’s alpha among comments and ratings were conducted. Cronbach’s alphas were nearly perfect across time points indicating significant convergence and suggesting redundancy among the 10 positive comments and seven “Personal and Social Development” items. The 10 positive comments fell in the nearly perfect range for both marking period one ($\alpha = .96$) and marking period four ($\alpha = .97$). The “Personal and Social Development” items also fell in the nearly perfect range for both marking period one ($\alpha = .91$) and marking period four ($\alpha = .95$).

The use of total scores for comments and rating items in regression analyses was supported by these findings. The 10 positive comments were summed across academic subjects

and all four marking periods to produce a total score ($\alpha = .98$.) The seven “Personal and Social Development” items were combined to provide a total score by summing ratings on all seven items for each of the four marking periods ($\alpha = .94$). The use of total scores for typical comments and “Personal and Social Development” items was additionally appropriate for this study as these scores provided a summative score for the full academic year across subject areas and marking periods, similar to a cumulative GPA. Cumulative GPA was utilized in regression analyses discussed below and the development of these total scores provided a comparable set of scores (i.e., representative of cumulative performance) for assessing potential associations.

To promote more thorough examination of the second research question and the Likert rating approach to SEL assessment on report cards, conceptual categorizations of “Personal and Social Development” items based on the CASEL 5 served as an additional data reduction strategy (i.e., Self-Awareness and Management, Social-Awareness and Relationship Skills, Responsible Decision-Making). Two additional summary scores were utilized to represent SEL proficiency and deficiency. The SEL proficiency total score was based on the frequency of check plus ratings given to each individual student across the four marking periods. The SEL deficiency total score was based on the frequency of check minus ratings given to each individual student across the four marking periods. This will allow for specific investigation of the any additional variance attributable to the most and least consistent demonstrations of SEL related behaviors in predicting cumulative GPA and performance on the NJASK.

Table 11 provides a correlation matrix for variables considered in data reduction decisions. Descriptive statistics and mean comparisons via Independent Sample T-tests were utilized to examine differences in achievement variables (e.g., NJASK scores, GPA) variables between genders. No significant gender differences were found for any of the achievement

variables. These results and descriptive statistics for achievement variables are provided in Table 12. Descriptive statistics for variables utilized in Regression Analyses are depicted in Table 13.

RQ2 Concurrent Validity

The two hypotheses for this question were tested via hierarchical multiple regression analyses. This method is most appropriate when theory testing as the stepwise method is susceptible to influence by random variation in the data which decreases the probability of replication (Field, 2009). Overall, this method was chosen as it produced the most parsimonious and conservative interpretation.

Issues of multicollinearity among rating items prevented meaningful analysis of the relationship between each of the three conceptual SEL skill categories and achievement as measured by GPA and NJASK scores. This issue was anticipated after findings indicating the nearly perfect internal consistency between rating items. See Table 11 which, as discussed above, presents a correlation matrix demonstrating the relationship between achievement and comment variables considered in the data reduction decisions.

H2a: Positive Comments and Higher Ratings Associated with Higher GPA. Hierarchical multiple regression analyses were performed with cumulative GPA being the dependent variable (DV). Step one included gender, step two included the total score for positive comments, and step three included the total score for “Personal and Social Development” items (PSD total). The overall model was able to explain 43% of the total variance in cumulative GPA. Gender was found to not explain a significant portion of the total variance. The positive comment total score was associated with cumulative GPA, with a large ES for unique variance explained, and the PSD total was also associated with cumulative GPA with a small ES, accounting approximately for an additional 9% of the variance. The ΔR^2 and β values are presented in Table 14.

The analysis was then replicated using two dependent variables: (1) Cumulative GPA for reading and writing, (2) Cumulative GPA for math and science. For the academic subject areas of reading and writing, the overall model again accounted for approximately 43% of the total variance. The positive comment total was associated with GPA for reading and writing with a large ES for unique variance explained. The PSD total was also associated with GPA for reading and writing with a small ES, again accounting for approximately 9% of the total variance explained (see Table 15). For the academic subject areas of math and science, the overall model accounted for approximately 37% of the total variance. The positive comment total associated with GPA for science and math with a large ES for unique variance explained, and the PSD total was also associated with GPA for science and math with a small ES, again accounting for approximately an additional 9% of the total variance explained (See Table 16).

H2a: Positive Comments and Higher Ratings Associated with Higher NJASK.

Hierarchical multiple regressions were conducted with NJASK performance in Language Arts and Math each serving as a dependent variable in separate analyses. Step one included gender, step two included the total score for positive comments, and step three included the PSD total. The overall model explained 23% of the total variance among scores on the Language Arts NJASK (see Table 17). For Math NJASK scores, the overall model explained 28% of the total variance (see Table 18). Gender was found to not account for a significant portion of the total variance explained for either DV. For both Language Arts and Math NJASK scores, the positive comment total score was associated with a medium ES for unique variance explained, and the PSD total was associated with a small ES for unique variance explained, accounting for approximately an additional 8% of additional variance. The ΔR^2 and β values are presented in Table 17 for the Language Arts NJASK analysis and in Table 18 for the Math NJASK analysis.

H3a: Comments and Ratings Related to Social Awareness and Relationship Skills Associated with Achievement. Issues of multicollinearity discussed above prevented meaningful analyses of the relationship between each conceptual categorization of comments and rating items and achievement. In response, two total scores representing SEL proficiency and SEL deficiency were created in order to better investigate the significant percent of variance in achievement (i.e., GPA and NJASK scores across subject areas) consistently accounted for by the “Personal and Social Development” items total score.

With reading and writing cumulative GPA and science and math cumulative GPA as the DVs, step one included gender, step two included the total score for positive comments, step three included the total score for SEL proficiency, step four included the total score for SEL deficiency, and step 5 included the total score for PSD. The overall model explained 44% of the total variance in reading and writing cumulative GPA, and 38% of the total variance in science and math cumulative GPA. In both analyses, positive comments, consistent with previous analyses for GPA, were significantly associated with GPA with a large ES for unique variance explained. For reading and writing, SEL proficiency and deficiency were significantly associated with GPA with a small ES for unique variance explained while respectively accounting for 6% and 4% of the total variance (see Table 19). For science and math, SEL proficiency and deficiency was significantly associated with GPA with a small ES for unique variance explained while respectively accounting for 5% and 6% of the total variance (see Table 20). The PSD total score did not account for a significant portion of variance after SEL proficiency and deficiency scores were considered in either analysis.

Altering the order of steps by adding the SEL deficiency score first and then the SEL proficiency score yielded similar results. Both were significantly associated with reading and

writing GPA with a small ES for unique variance explained with deficiency accounting for 7% and proficiency 3% of the total variance (see Table 21). For science and math both deficiency and proficiency were significantly associated with GPA, however, deficiency accounted for 10% of the total variance explained and proficiency was found to no longer be significant (see Table 22).

With Language Arts NJASK scores as the DV, analyses were replicated. The overall model explained 20% of the total variance. SEL proficiency scores were significantly associated with Language Arts NJASK scores with small ES for unique variance explained. SEL deficiency scores, however, were not significant (see Table 23). When deficiency was entered and then proficiency, both were significantly associated with Language Arts NJASK scores with a small ES for unique variance explained with approximately 4% of additional variance accounted for by each (see Table 25). Analyses were again replicated with Math NJASK scores as the DV. When SEL proficiency scores were entered first, both proficiency and deficiency scores were significantly associated with Math NJASK scores with a small ES for unique variance explained with approximately 5% of total variance explained accounted for by each (see Table 24). When SEL deficiency scores were entered first, deficiency scores were significantly associated with Math NJASK scores with a small ES for unique variance explained, accounting for approximately 8% of the total variance explained. Proficiency scores, however, were not significantly associated with Math NJASK scores after accounting for SEL deficiency scores (see Table 26).

H3a: Accounting for Previous Achievement in Association Between Comments, Ratings, and Achievement. To further investigate the concurrent validity of positive comments and PSD ratings for academic achievement, additional hierarchical multiple regressions that

included previous achievement were conducted. Data available were limited to a single academic year, and thus cumulative marking period one GPA for reading and writing (RW-MP1GPA) and science and math (SM-MP1GPA) were selected as the most appropriate independent variables of previous achievement. In selecting the DVs for achievement, issues of multicollinearity prevented the use of cumulative GPA for the respective subject areas, as was done in previous analyses, as RW-MP1GPA and SM-MP1GPA contributed to that total score. Marking period four GPA for the respective subject areas (RW-MP4GPA, SM-MP4GPA) was selected as the DV as it provided the best source of data to compare with MP1GPA longitudinally.

The initial analysis utilized RW-GPAMP4 as the DV. Step one included gender, step two included the RW-MP1GPA, step three included the total score for positive comments, and step four included the PSD total. The overall model explained 57% of the total variance for RW-MP4GPA. Gender was found to not account for a significant portion of the total variance explained. RW-MP1GPA was significantly associated with RW-GPAMP4 and accounted for approximately 54% (medium ES) of the variance explained. The positive comment total did not account for a significant portion of the variance explained. The PSD total score was significantly associated with RW-MP4GPA with a small ES for unique variance explained, accounting for approximately an additional 3% of variance. The ΔR^2 and β values are presented in Table 27.

The next analysis utilized SM-GPAMP4 as the DV. Step one included gender, step two the SM-MP1GPA, step three the total score for positive comments, and step four the PSD total. The overall model explained 62% of the total variance for SM-MP4GPA. Gender was found to not account for a significant portion of the total variance explained. SM-MP1GPA was significantly associated with SM-MP4GPA and accounted for approximately 56% of the variance (medium ES). The positive comment total and the PSD total score were significantly

associated with SM-MP4GPA with a small ES for unique variance explained accounting for approximately 4% and 3% of the total variance explained respectively. The ΔR^2 and β values are presented in Table 28.

Chapter IV

Discussion

This study aimed to further investigate how report card comment sections relate to academic achievement, through an SEL lens. This was the first examination of both typical report card comments and Likert ratings of specific behaviors based on early SEL skills theory in the report card comment section. Analyses of 113 report cards from a suburban elementary school revealed that 19 of the 24 comments were qualitatively related to the skills described by the CASEL 5; however, 14 of the 24 typical report card comments were assigned less than one time on average across all marking periods and subject areas, indicating little utility. The 10 remaining comments and “Personal and Social Development” (PSD) Likert items were qualitatively related to the CASEL 5 skills. The 10 positive comments initially significantly associated with cumulative GPA across subject areas with a large effect size were found to retain a small effect size on GPA in science and math after accounting for previous GPA. However, comments were not significantly associated with cumulative GPA for reading and writing after accounting for previous GPA. Positive comments were significantly associated with standardized test scores with a medium effect size. The PSD score was a significant predictor of NJASK scores after accounting for the effects of positive comments. The PSD score was also a significant predictor of GPA across subject areas after accounting for the effects of previous GPA. Results from regression models showed that PSD deficiency scores explained greater variance than PSD proficiency scores for math and science GPA and math scores on the NJASK.

Comparison with Existing Literature

Findings from the current study are consistent with emerging evidence that students who received more positive SEL related report card comments, in this study represented by the ten

positive comments, had higher letter grades and better performance on standardized tests in both “subjective” (i.e., language) and “objective” (i.e., math) academic areas (Moceri et al., 2014 & Moceri & Elias, 2014). Similar to other studies in this area was that comments were only partially able to assess the five core aspects of SEL (i.e., the CASEL 5). Specifically, in this study, the CASEL 5 skill area of Responsible Decision-Making was not represented by typical comments. Additionally, it is important to note there are a broad range of skills in each of the CASEL 5 skills areas (see Appendix A), and thus, while comments related to Self-Awareness, Self-Management, Social-Awareness, and Relationship Skills were present, only a few specific skills from each area were represented by comments (discussed further below). It is clear that to systematically assess the impact of all areas of SEL skill, items covering multiple skills from each of the CASEL 5 skill areas must be intentionally included in comment sections.

Also consistent with previous research is the underutilization of many comments provided in a menu format to teachers (Friedman et al., 1998). Friedman et al. (1998) examined a computerized drop down menu of over 80 comment options and found that teachers tended to use a total of 17 comments options for all comment assignments, and that, despite no limitation, approximately half of the students in the study received two or fewer comments. Moceri et al. (2014) attributed infrequently assigned comment options to restrictions placed on teachers to only provide two or three comments per student. The current study, however, demonstrated that similar to Friedman et al. (1998), with a menu of 24 comments provided directly on the report card itself, 14 of those items on average were assigned to students less than once across all marking periods and academic subjects. These findings converge in demonstrating the lack of practical utility of many typical report card comment systems.

While individual student data regarding SES and ethnicity were not available to compare with previous research documenting discrepancies in comment assignment based on these variables, gender was not found to be associated with such discrepancies. This differs from the findings of Kemp et al. (2014) and Mocerri et al. (2014), which showed males tended to have worse academic outcomes (i.e., more negative comments, fewer positive comments, lower letter grades, and lower standardized test scores). It is important to note that the elementary student population in this study did not demonstrate significant differences in GPA or performance on the NJASK based on gender. This is particularly important given findings in this study that comments were significantly associated with GPA to the extent that when prior GPA was accounted for, typical comments did not account for a significant portion of unique variance explained in predicting academic achievement. Thus, it may be expected that comment assignment did not significantly differ for male and female students given that comments were likely reflective of the overall consistency in academic achievement across genders.

The lack of discrepancies between genders in academic achievement and, relatedly, in comment assignment may be further explained by the age of the population in this study. In previous research that supported gender based discrepancies, Kemp et al. (2014) examined middle school students and Mocerri et al. (2014) studied students at this high school level. While some research supports that females demonstrate higher academic achievement consistently across grade levels (Voyer & Voyer, 2014), other data suggests that discrepancies in achievement become more significant as students ascend into middle school (i.e., grades and above) (Zembar & Blume, 2009). Thus, it might be that the elementary students in this study were more likely to demonstrate consistency in achievement and comments across genders than those examined at higher grade levels. Additionally, demographic variables have been found to

be a significant moderator in gender discrepancies related to achievement (Voyer & Voyer, 2014) suggesting that had more comprehensive demographic data been available for this study, nuances related to gender, demographics such as ethnicity, and achievement may have emerged.

Distinct for this study was the examination of a Likert rating system based on SEL theory (i.e., “Personal and Social Development” items or PSD). This provided additional examination of behavioral items developed based on SEL programming and supported previous comment related findings that the presence or absence of SEL skills, and more specific to this study the frequent or infrequent demonstration of SEL skills, was associated with higher and lower academic grades and standardized test scores. This study’s findings that Likert rating items were significant in accounting for unique variance explained beyond what was accounted for by typical comments suggests Likert items had stronger concurrent validity for achievement. Sensitivity to the frequency with which these skills were demonstrated through the three-point Likert rating system, as opposed to the dichotomous comment variables which can only speak to the presence or absence of a behavior, provided additional information likely essential to bolstering validity.

While with the intentional design of SEL into the Likert rating system included coverage of the three overarching skill categories of the CASEL 5 (i.e., Self-Awareness and Self-Management, Social-Awareness and Relationship Skills, and Responsible Decision-Making), not all domains of the CASEL 5 were equally represented. Specifically, Self-Awareness and Self-Management were represented by three items, Social-Awareness and Relationship Skills were represented by three items, and Responsible Decision-Making was represented by one item. In examining Likert rating items with greater detail, the two skill categories regarding “awareness” (i.e., Self-Awareness and Social Awareness) were not well represented. Items in each of the three

overarching skill categories focused on the demonstration of a skill contingent on awareness, such as using skills to regulate emotions, rather than having items specifically dedicated to the recognition of specific emotions or self-identification of the need to use regulatory strategies. Thus, the skill areas of Self-Management and Relationship skills were better represented. Similarly, with the area of Responsible Decision-Making being represented by only one item (i.e., “identifies and accepts responsibility for actions”), skills in this area typically reflective of internal processes (e.g., identifying pros and cons) were not addressed. It may be that these skills are less readily observable, particularly without specific interventions or activities designed to allow these skills to be demonstrated.

Considerations and Potential Explanations for Current Findings

Qualitative categorization of typical report card comments onto three overarching categories of the CASEL 5 (i.e., self-awareness and management, social awareness and relationship skills, responsible decision making) identified 19 items as related to the CASEL 5. This can partially be explained by the emergent literature and policy that has supported the promotion of SEL and related skills in schools (O’Connell, 2009) and the diverse array of theories and fields from which SEL and related skills emerged (Watson & Emory, 2010; Wigelsworth et al., 2010). Consider first that SEL emerged from a variety of fields and theoretical perspectives, thus creating a broad umbrella of potential behaviors representative of the varying conceptualizations of SEL. Scholars at organizations such as CASEL systematically review literature on a large scale in order to be comprehensive in the investigation of behaviors potentially representative of each CASEL 5 skill category. This process has led a broad spectrum of behaviors to be included in definitions and lists of skills related to SEL thus increasing the chance that behaviors on report cards would be encompassed.

In addition to the breadth of behaviors now identified as related to SEL due to the diverse body of literature that led to its development, practice, policy, and research regarding SEL in educational settings has continued to expand the development of specific operationalized definitions of SEL skills in schools. Behaviors specific to educational settings have been articulated with increasing frequency. These specific definitions encompass common observable student behaviors to assist educators in identifying when a behavior is representative of an SEL skill. Thus, when current skill definitions are reviewed, such as those used for this study (see Appendix A), many behaviors commonly seen in schools are included. State standards (e.g., Indiana, Pennsylvania) where SEL behaviors and skills have been systematically infused into academic lessons and routines serve as a strong example of how this focus on student behaviors related to SEL has made the inclusion of such behaviors in definitions and skill examples more widespread. These efforts in combination lead many behaviors described by typical comments to be readily viewed through an SEL lens as is done in systematic research, policy, and programming.

The infrequency with which 14 of the 24 comments were assigned may best be explained by previous research in this area (e.g., Mocerri et al, 2014, Mocerri & Elias, 2014). These studies identified teachers as more likely to assign positive comments, even when more negative than positive comment options are available. In the present study, positive and negative comments largely served as mirror images reflecting the presence (i.e., positive) or absence (i.e., negative) of the same behavior. There were an even number of positive and negative comment options, and given that many of the negative comments did not address behaviors outside the scope of positive comments, the redundancy likely negatively impacted a more equal distribution of comment assignment. The lack of nuance between many of these comments, even among

positive items, was demonstrated by the significant intercorrelations between all 10 of the most frequently given positive comments.

Interestingly, the seven “Personal and Social Development” items also yielded one factor models and high correlations between all items. One potential explanation is that the SEL programming from which the district derived these items was not as deliberate as the CASEL 5 in differentiating between skill groupings. Similarly, it might be that educators in the district were not deliberately sensitive to differentiating skills groupings in the process of developing items. Questions regarding teacher training on how to recognize different skill areas and provide ratings might also serve to explain these findings. This might be particularly so as, from a face validity standpoint, teachers likely observed natural connections between each of these behaviors. Connections between items may have been reinforced as all items were listed under a single scale on the report card itself, potentially making it more likely that ratings across items would be similar.

Further explanations for the PSD items loading onto a single factor may be gleaned from issues in the field of SEL assessment in general related to limited specificity (McKown, 2015, Denham, 2015) and comprehensiveness (Wigelsworth et al., 2010). Many measures of SEL do not address all aspects of the CASEL 5, and almost none do so in a comprehensive enough manner for skill development between areas to be accurately compared (Denham et al., 2010; Haggerty, Elgin, & Woolley, 2011). As a result, detailed examinations differentiating skill development trajectories for each of the CASEL 5 skill areas were lacking in the research at the time of this study.

Two potential explanations for PSD items loading onto a single factor stem from the above regarding the state of SEL research. One explanation is rooted in the need to clearly

articulate a developmental trajectory for the CASEL 5 skills and design rating systems accordingly. One of the few examples of such a trajectory is the SEL developmental standards articulated by the public school district in Anchorage, Alaska where indicators have the necessary specificity for the school age developmental continuum, but have yet to be operationalized into report cards (<http://alaskaice.org/school-climate/anchorage/>). A second explanation is based on how the recent and not yet widespread development of such a trajectory has led to a current lack of integration into assessment. This has negatively impacted the developmental comprehensiveness and sensitivity of even more rigorous measures, and limited investigations aimed to differentiate between skill areas by teams of researchers and experts in the field. Thus, it is likely that this brief seven item Likert rating system developed by the school district lacked the specificity, comprehensiveness, and psychometric properties to sensitively measure discrepancies in development across skill areas.

Results pertaining to typical comments from regression analyses may be explained by the following. Typical report card comments originated historically as feedback on behaviors reflecting the school culture's emphasis on academic grades and the behaviors that appear to support achievement (e.g., brings materials to class). Given the paucity of research on report card comments (Moceri et al., 2014), particularly before the work of the SEL Lab at Rutgers University, the validity of these "academic" behaviors lacked examination and comments, instead, emerged through practice as a supplemental form of information to support academic grades. This is demonstrated by Moceri et al. (2014) through findings that negative comments tended to be designated in order to explain poor grades. The historical relationship between academic grades and comments has become routine practice, and thus, comment assignments are likely to closely follow academic grades (GPA). The close development of comments to support

academic grades rather than speak to behaviors that support general achievement was demonstrated in current findings with smaller effect sizes for unique variance accounted for by typical comments in regards to standardized assessments than with academic grades. This is further demonstrated through findings that when previous achievement was accounted for, even typical comments' relationship with academic grades was non-significant or reduced from medium or large to small effect sizes for unique variance explained.

Personal and Social Development (PSD) items consistently demonstrated a significant association with achievement across academic subjects and standardized test scores with a small effect size for unique variance explained. This is consistent with research demonstrating that SEL skills can significantly impact academic achievement (e.g., Durlak et al., 2011), particularly given that these items were based specifically on the sample school's SEL programming and thus demonstrated a clearer and more deliberate focus on general SEL theory than typical comments. This may help to explain how PSD items maintained a significant relationship and consistent effect size for unique variance explained across multiple forms of achievement. SEL theory and the accompanying skills emerged from a variety of fields and theories with an initial deliberate focus on promoting the well-being and success of children in life rather than specifically in academics. Thus, the PSD items were more reflective of an emphasis on skills that promote general success when compared to typical comments that dominantly addressed "academic behaviors" rooted in grades.

Also, this emphasis may have encouraged greater independence of PSD item ratings from academic grade assignment. Consider how items addressing how to manage one's self, interact with others, and make responsible decisions in a highly social school context may be more likely to deviate from grade assignment than those behaviors that emerged alongside academic grades

and have demonstrated use as an attempt to explain academic grades (Moceri et al., 2014). This likely promoted recognition from raters that these items would not serve the same function of “explaining” poor grades, but rather speaking to these specific observable behaviors rooted in SEL theory. This differentiation likely allowed PSD items to maintain a relationship with academic grades without being persistently or rigidly offered in combination with a particular grade.

Another explanation of the PSD items significant relationship with multiple forms of academic achievement may stem from the additional variability present for Likert items. Additional variability was present in PSD total scores and PSD proficiency and deficiency totals in comparison to the dichotomous comment variables. This likely promoted greater sensitivity in measurement which was demonstrated by statistically significant relationships with academic grades and standardized tests with small effect sizes for unique variance explained after the model accounted for the proportion of total variance explained attributable to typical comments. Findings from PSD proficiency and deficiency totals showed that the range of check plus and check minus rating options were used frequently enough to produce statistically significant relationships and effect sizes for unique variance explained similar in value to what the PSD total scores produced. This indicates that when given the option, teachers used a range of three options to address the frequency with which skills are exhibited effectively enough to predict achievement beyond merely indicating if a behavior was present or not. These findings demonstrated that shifting away from a system of dichotomous variables can produce more varied data better suited for maintaining significant relationships with achievement in more sophisticated analyses.

Findings related to PSD scores raised questions of how relevant the presence as opposed to absence of specific skills may be for academic outcomes. For both NJASK Math scores and science and math combined cumulative GPA, PSD deficiency was found to have a significant relationship with a small effect size and proficiency was not significantly related. This may indicate that absence or particularly infrequent demonstration of SEL skills is more significant than the presence or frequent demonstration. Stated differently, it may be that the harm of not having SEL skills is more significant than the benefit of consistent skill demonstration.

Limitations & Suggestions for Future research

Multiple limitations are important to consider and should help inform future research. First, limitations in the data prevented the use of multi-level modeling and the inclusion of class and teacher levels in the analyses. In many cases in this study, the same teacher provided letter grades, typical comments from the list of 24, and ratings on “Personal and Social Development” items. While clear data were provided for some students as to which teacher provided which letter grade, comment, and/or rating, approximately over 40 students were assigned comments by two teachers with no clear indication of which teacher provided which letter grade, comment, and/or rating. This is problematic as statistical dependence can occur when the same individuals (teachers) provide ratings on the same persons (students) using multiple dimensions. In order to maintain the assumption involved in many regression strategies, such as the one used in this study, that each individual provides a unique piece of statistical information unrelated to the information provided by other students in the sample, multi-level modeling is necessary. Without a clear indication of which teachers for what classes provided which data for the entirety of the sample, it was not possible to utilize multi-level modeling that included class and teacher levels in the analysis thus leaving the data susceptible statistical dependence. This may have

negatively impacted the validity of conclusions drawn from the analysis as the use of multi-level modeling to account for statistical dependence promotes unbiased estimates of standard errors associated with regression coefficients and increases the accuracy of statistical conclusions (O'Dwyer & Parker, 2014).

Another related concern is that ratings related to SEL (i.e., PSD items) and GPA are conflated given mono-informant (i.e., teacher) bias. Specifically, teacher ratings on PSD items are likely interactively connected to the assignment of academic grades. While restricted in data, this concern regarding shared variance was addressed in the present study by determining the strength of the relationship between cumulative GPA and NJASK scores, positive comments and NJASK scores, and PSD total and NJASK scores via Pearson's correlations. Findings yielded very large to nearly perfect correlations, mitigating the potential influence of teacher bias given that across raters, student GPA, positive comments and the PSD total maintained a strong relationship with an objective assessment of academic achievement; NJASK scores.

Future data collection should support multi-level modeling by including which teacher(s) provided which comments or ratings for every student and connectedly the classes in which teachers provided ratings for each student. Several studies show support for the relationship between academic grades, demographics, and report card comments and ratings, indicating the stability of that relationship. However, the use of multi-level models, such as hierarchical linear modeling, will promote examination and development of SEL assessment in a manner that differentiates individual level variance from the shared variance of students being rated by the same teacher, and as a result more accurately pinpoints the relationship between SEL comments/ratings and academic grades as appropriate for each individual student.

Demographic data for each individual student and teacher should be collected in future studies and included in multi-level modeling. The findings of both Kemp et al. (2014) and Moceris et al. (2014) suggest that individual student characteristics such as ethnicity, gender, and SES may have a significant relationship with report card comment assignment trends and academic achievement. Presently, the role of demographic variables is unclear and requires considerably more examination in order for findings to be generalizable. In this study, the lack of demographic data beyond gender prevented analyses and interpretations from being conducted in a culturally sensitive manner able to examine subgroup discrepancies and consistencies. The role of teacher demographic variables on report card comment sections has yet to be examined and thus data for both the students and teachers in future samples should be collected and interaction effects investigated.

Examining demographic data may provide valuable information regarding how gender and culture-based differences impact the rating and demonstration of student behaviors related to SEL. Hofmann (2006) highlights the importance of culture by referring to research showing differences in how individuals from different cultures have different values and beliefs (e.g., spirituality, sense of self), styles of thinking (e.g., focusing on environment/context as whole or individual characteristics), and methods of reasoning (e.g., using analytical logic to resolve to contradicting ideas as opposed to identifying and accepting the contradiction as a dialectic). Elias (2015) highlights how the development of multicultural perspectives can be bolstered by, and to some extent requires, SEL skills related to understanding one's self, communicating effectively to understand the feelings and opinions of others, and considering decisions with sensitivity to context and social norms.

Similarly, the importance of culture in influencing beliefs, thinking, and reasoning (Hofmann, 2006) should also be considered with regard to students' caregivers. A potential advantage of providing feedback related to SEL on a report card is the opportunity to make conferences between teachers and caregivers more meaningful, and for these skills to be reinforced and modeled by caregivers at home. This may promote skill development, as well as open communication between parents, teachers, and students regarding behaviors important to promoting student success. If culture is not considered in a manner that allows for open communication and understanding of SEL skills and their importance through appropriately tailored discussions and materials, a valuable opportunity to engage caregivers in student SEL skill development may be missed. Additionally, caregivers may misunderstand the purpose of providing feedback on these skills, potentially leading to conflict or opposition regarding SEL rating systems in schools. This may be particularly problematic, given that a deliberate focus on SEL skills on a large scale in schools is a relatively recent initiative and clear policy at the federal level has yet to explicitly address the instruction of SEL and related fields in schools. This creates opportunity for parents, for example, whose cultural norms view the school as focusing only on academics and do not see SEL as purview of schools, to resist the promotion of these important skills and thereby inadvertently increase conflict between parents and educators.

From a measurement standpoint, another important limitation of this study is inherent in the nature of the way report card comments are typically configured nation-wide. They are dichotomous and many are assigned infrequently. This severely limits the psychometrics that can be applied and the amount of information about students' SEL and non-SEL skills and behaviors that can be gleaned from typical comments. The dichotomous nature of the typical comment variables (i.e., absent or present) lacks sensitivity to levels of skill proficiency and frequency of

demonstration thus restricting the range of responses. These issues prevent most school systems from systematically using report card comments to monitor program effectiveness or even to track behavior trends overall or by subgroups. In this study, it was not possible to include all twenty-four comments in factor analyses and thus, comment groupings by factor were unable to include all comments in conceptual interpretation (e.g., consistent or inconsistent with the CASEL 5). This is of particular significance given that the ten comments for which there was sufficient data were found to load onto a single factor. It might be that had the other comments had sufficient data to be included in the analysis, additional factors may have emerged. Mocerri et al. (2014) encountered a similar limitation; however, it was largely attributed to teachers in that district being only permitted to provide two comments per student. This study highlights that even without restricting the number of comments assigned to each student teachers presented with a list of twenty-four comments only consistently assigned a limited number ($m = 3.86$).

Future research should additionally collect data so as to provide a more systematic baseline of academic achievement, preferably through both academic grades and standardized tests from prior years. This is supported by this study, as well as Mocerri et al.'s (2014) finding that the predictive power of report card comments was reduced significantly when controlling for previous academic grades. Including baseline data will allow for more accurate investigations of causal links between SEL comments/ratings and achievement beyond what previous achievement would suggest and with greater sensitivity to potential confounds. Additionally, baseline data on both academic achievement and SEL comments/ratings would allow for longitudinal investigation of the relationship including potential growth trajectories for SEL in relation to academic achievement as students ascend developmentally through grade levels. That said, this does not address the difficulty of disentangling the overlap in observations about the

comments also being part of the judgment going into the formulation of the report card academic grades.

Future research is likely best focused on comment systems that are intentionally focused on assessing SEL and related skills. Then, well-established brief measures of SEL (e.g., Devereux Student Strengths Assessment Mini, an 8-item screening tool) may serve both as a model for brief rating sections and as a comparative tool through which convergent validity can be established.

Systematic data collection including the potential use of well-established measures of SEL should additionally consider data analytic strategies to further examine the relationships of SEL proficiency and deficiency and academic achievement. Negative comments were previously found to be given significantly less often than positive comments, and to demonstrate a significant relationship with achievement despite low frequency (Moceri et al., 2014). Similarly, the present study demonstrated that SEL deficiency may serve as a more robust predictor of achievement than SEL proficiency. However, the present findings should be investigated further to determine if tracking students' deficient SEL ratings may serve as a universal screening tool and assist in identifying those students in need of intervention. In general, the potential for SEL ratings to identify students most in need socially, emotionally, behaviorally, and/or academically may provide an additional benefit to assessing SEL universally.

As policy leads schools and districts to a more deliberate SEL focus, overall school climate should also be assessed. The general school environment influences the individual students' SEL skill development (Elias, 2009). Teacher ratings of SEL skills may change as the norms within the school lead to changes in expected student behavior.

Implications and Suggestions for Current Practice

Schools should reevaluate report cards comments as currently used in practice and replace or alter comment content and focus. Comments that are unhelpful or not actionable, meaning that the comment does not readily lend itself to a specific skill or behavior to promote, should be addressed. Analyzing the frequency with which comments are offered may serve to highlight comments of little practical value. Previous research reviewed for this study indicated that ambiguous comments that do not address a specific behavior with an underlying skill are misleading as they appear positive, but follow the same pattern of assignment as negative comments. Thus, these comments do not provide additional information or an accurate indication of positive change (Moceri et al., 2014).

Comments focused on skills related to SEL, however, can highlight important areas capable of being taught and reinforced by both educators and parents. This is of particular importance, given clear research demonstrating as much improvement in academic achievement as purely academic interventions when skills related to SEL are fostered (CASEL, 2005; Durlak et al., 2011).

Revising report card comment structure is another important consideration. The elementary school in this study serves as an example of how a brief set of Likert rating items applied uniformly to all students can increase measurement sensitivity to these critical skills, readily allow for comparison of rating items and typical comments, and provide stronger data to inform report card based measure development. Having a common set of ratings for all students is essential for longitudinal evaluation and for schools to be able to derive policy implications from the other side of the report card. For schools both with and without existing SEL programming, developing a ratings system based on SEL or related skills is feasible and can be

supported through acquisition of resources developed specifically for the purpose of adapting report card comment sections (e.g., Ferrito, Elias, & Mocerri, 2015). Further, it is likely that as schools come to highlight the presence of SEL skills on their report cards, they will also tend to increase intentional efforts toward developing those skills in their students.

References

- Association for Supervision and Curriculum Development (2007). *The learning compact redefined: A call to action—A report of the Commission on the Whole Child*. Alexandria, VA: Author. Last retrieved March, 21 2015 from <http://www.ascd.org/ASCD/pdf/Whole%20Child/WCC%20Learning%20Compact.pdf>
- Bernard, M.E., Elias, M.J., Bell, P. J., Ferrito, J. J. & Langione, K. (2015). Social and Emotional Learning: Role of Psychologists in Australia. In M. Thielking & M. Terjesen (Eds). *Handbook of Australian School Psychology: International Research, Practice and Policy*. New York: Springer.
- Christ, T.J., Riley-Tillman, T.C., Chafouleas, S., and Jaffery, R. (2011). Direct behavior ratings: An evaluation of alternate definitions to assess classroom behaviors. *School Psychology Review, 40*(2), 181-199.
- Cohen, J. (1992). A Power Primer. *Psychological 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 (2015a). *Federal legislation to promote social and emotional learning*. Retrieved from <https://casel.squarespace.com/federal-policy-and-legislation/>
- Collaborative for Academic, Social, and Emotional Learning (2015b). *Social and emotional learning core competencies*. Retrieved from <http://www.casel.org/social-and-emotional-learning/core-competencies/>
- Collaborative for Academic, Social, and Emotional Learning (2015c). *What is social-emotional learning?* Retrieved from <http://www.casel.org/social-and-emotional-learning/>

- Conley, D.T. (2014). New conceptions of college and career ready: A profile approach to admission. *Journal of College Admission*, 223, 12-23.
- Crowe, L.M., Beauchamp, M.H., Catroppa, C., & Anderson, V. (2011). Social function assessment tools for children and adolescents: A systematic review from 1988 to 2010. *Clinical Psychology Review*, 31, 767-785.
- Denham, S.A. (2005a). *Assessing social-emotional development in children from a longitudinal perspective for the National Children's Study: Social-emotional compendium of measures*. Columbus, OH: Battelle Memorial Institute.
- Denham, S.A. (2005b). *Assessing social-emotional development in children from a longitudinal perspective for the National Children's Study*. Columbus, OH: Battelle Memorial Institute.
- Denham, S.A. (2015). Assessment of SEL in Educational Contexts. In J.A. Durlak, C.E. Domitrovich, R.P. Weissberg, & T.P. Gullotta (Eds.), *Handbook of Social and Emotional Learning* (285-300). New York & London: The Guilford Press.
- Denham, S.A. & Weissberg, R.P. (2004). Social-emotional learning in early childhood: What we know and where to go from here. In E. Chesebrough, P., King, T.P., Gullotta, & M. Bloom (Eds.), *A blueprint for the promotion of prosocial behavior in early childhood* (13-50). New York: Kluwer Academic/Plenum Publishers.
- Denham, S.A., Ji, P., & Hamre, B. (2010). *Compendium of preschool through elementary school social-emotional learning and associated assessment measures*. Retrieved from <http://www.casel.org/library/2013/11/4/compendium-of-preschool-through-elementary-school-social-emotional-learning-and-associated-assessment-measures>

- Durlak, J.A., Domitrovich, C.E., Weissberg, R.P. & Gullotta, T.P. (Eds.). (2015). *Handbook of Social and Emotional Learning* (285-300). New York & London: The Guilford Press.
- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D. & Schellinger, K. B. (2011), The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, 82: 405–432.
- Dusenbury, L. (2014). [What are the key features of high-quality standards for SEL?](#) Chicago: Collaborative for Academic, Social, and Emotional Learning.
- Dusenbury, L., Weissberg, R. P., Goren, P., & Domitrovich, C. (2014). State standards to advance social and emotional learning: Findings from state scan of social and emotional learning standards, preschool through high school, 2014. Chicago: Collaborative, for Academic, Social, and Emotional Learning.
- Edmunds, L. & Steward-Brown, S. (2003). *Assessing emotional and social competence in primary school and early years settings: A review of approaches, issues, and instruments*. University of Oxford: Health Services Research Unit Institute of Health Sciences.
- Elias, M.J. (2009). Social-emotional and character development and academics as a dual focus of educational policy. *Education Policy*, 23, 831-846.
- Elias, M.J. (2015). *How SEL can help students gain a multicultural perspective*. Retrieved from <http://www.edutopia.org/blog/how-sel-can-help-students-gain-multicultural-perspective-maurice-elias>
- Elias, M. J., & Bruene, L. (2005). Social Decision Making/Social Problem Solving: A Curriculum for Academic, Social, and Emotional Learning, Grades 4-5. Champaign, IL: Research Press

- 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.
- Elliot, S.N., Frey, J.R., & Davies, M. (2015). Systems for Assessing and Improving Students' Social Skills to Achieve Academic Competence. In J.A. Durlak, C.E. Domitrovich, R.P. Weissberg, & T.P. Gullotta (Eds.), *Handbook of Social and Emotional Learning* (301-319). New York & London: The Guilford Press.
- Field (2009). *Discovering Statistic Using SPSS* (3rd ed.). Thousand Oaks, CA: SAGE Publications Inc.
- Friedman, S. J., & Frisbee, 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.
- Gresham, F. M., Elliott, S. N., Cook, C. R., Vance, M. J., & Kettler, R. J. (2010). Cross-informant agreement for social and problem behavior ratings: An investigation of the Social Skills Improvement System—Rating Scales. *Psychological Assessment*, 22, 157–166.
- Haggerty, K., Elgin, J., & Woolley, A. (2011). *Social-Emotional Learning Assessment Measures for Middle School Youth*. Seattle: Social Development Research Group, University of Washington Commissioned by the Raikes Foundation.
- Hopkins, W. G. (2002). A scale of magnitudes for effect statistics. *A New View of Statistics*. Retrieved May 15, 2015, from <http://sportsci.org/resource/stats/effectmag.html>

- Hudley, C. (2001). The role of culture in prevention research. *Prevention & Treatment, 4*(1), 5, doi: 10.1037/1522-3736.4.1.45c
- Humphrey, N., Curran, A., Morris, E., Farrell, P., & Woods, K. (2007). Emotional intelligence and education: A critical review. *Educational Psychology, 27*, 233–252.
- Humphrey, N., Kalmbouka, A., Wigelsworth, M., Lendrum, A., Deighton, J., & Wolpert, M. (2011). Measures of social and emotional skills for children and young people: A systematic review. *Education and Psychological Measurement, 71*(4), 617-637. doi: 10.1177/0013164410382896
- Indiana Department of Education (2012). *Foundations to the Indiana Academic Standards for Young Children Birth to Age 5*. Retrieved from <http://www.doe.in.gov/sites/default/files/curriculum/indianafoundations-february-2012.pdf>
- Kenziora, K., Weissberg, R. P., Ji, P., & Dusenbury, L. A. (2011). [Strategies for social and emotional learning: Preschool and elementary grade student learning standards and assessment](#). Newton, MA: National Center for Mental Health Promotion and Youth Violence Prevention, Education Development Center, Inc.
- Linkins, M., Niemiec, R.M., Gillham, J., & Mayerson, D. (2014). Through the lens of strength: A framework for educating the heart. *The Journal of Positive Psychology: Dedicated to Furthering Research and Promoting Good Practice, 10* (1), 1-5 <http://dx.doi.org/10.1080/17439760.2014.888581>
- Lipton, M.E. & Nowicki, S. (2009). The social emotional learning framework (SELF): A guide for understanding brain-based social emotional learning impairments. *Journal of Developmental Processes, 23*(4), 99-115.

- Marzano, R.J. (2015). Using Formative Assessment with SEL Skills. In J.A. Durlak, C.E. Domitrovich, R.P. Weissberg, & T.P. Gullotta (Eds.), *Handbook of Social and Emotional Learning* (285-300). New York & London: The Guilford Press.
- McKown, C. (2015). Challenges and Opportunities in the Direct Assessment of Children's Social and Emotional Competencies. In J.A. Durlak, C.E. Domitrovich, R.P. Weissberg, & T.P. Gullotta (Eds.), *Handbook of Social and Emotional Learning* (285-300). New York & London: The Guilford Press.
- Meyers, L. S., Gamst, G., & Guarino, A. J. (2006). *Applied multivariate research: Design and interpretation*. Thousand Oaks, CA: Sage.
- Miyamoto, K., Huerta, M.C., & Kubacka, K. (2015). Fostering social and emotional skills for well-being and social progress. *European Journal of Education*. doi: 10.1111/ejed.12118
- Moceri, D. C., Elias, M. J., Fishman, D. B., Epstein, Y., & Selby, E. A. (2014). *Social-emotional Competencies and Academic Achievement in Diverse High School Youth*. Under review.
- Moceri, D. C. & Elias, M. J. (2014). *Grading the other side of the report card: How comments on an elementary school's report card are connected to current and future academic performance*. Under review.
- OECD (2015). *Skills for Social Progress The power of social and emotional skills*. Paris, France: OECD Publishing.
- O'Dwyer, L.M. & Parker, C.E. (2013) *A primer for analyzing nested data: multilevel modeling in SPSS using an example from a REL study*. Washington, DC: US. Department of Education, Institute for Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands.
- Retrieved from http://ies.ed.gov/ncee/edlabs/regions/northeast/pdf/REL_2015046.pdf

- Renk, K., & Phares, V. (2004). Cross-informant ratings of social competencies in children and adolescents. *Clinical Psychological Review, 24*, 239-254.
- Sattler, J.M. (2008). *Assessment of children: Cognitive foundations (5th ed.)*. La Mesa: Jerome M. Sattler, Publisher, Inc.
- Snow, C. E., & Van Hemel, S. B. (2008). *Early childhood assessment: Why, what, and how*. Washington, DC: National Academies Press.
- Stecher, B.M. & Hamilton, L.S. (2014). *Measuring hard-to-measure student competencies: A research and development plan*. Santa Monica, CA: RAND Co.
- Strive Together (2013a). *Beyond content: Incorporating social and emotional learning into the strive framework. Volume 1: Social and emotional competencies and their relationship to academic achievement*. Retrieved from <http://www.strivetogether.org/resources/reports>
- Strive Together (2013b). *Beyond content: Incorporating social and emotional learning into the strive framework. Volume 2: A summary of measures by competency and stage of the cradle to career continuum*. Retrieved from <http://www.strivetogether.org/resources/reports>
- Strive Together (2013c). *Beyond content: Incorporating social and emotional learning into the strive framework. Volume 3: A compendium of social and emotional competency measures*. Retrieved from <http://www.strivetogether.org/resources/reports>
- Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140*(4), 1174-1204. doi:10.1037/a0036620
- Watson, D.L., & Emory, C. (2010). From rhetoric to reality: The problematic nature and assessment of children and young people's social and emotional learning. *British Educational Research Journal, 36*(5), 767-786.

- Weare, K., & Gray, G. (2003). *What works in developing children's emotional and social competence and wellbeing?* Research Report No 456. Crown Printers: University of Southampton, Health Education Unit, Research and Graduate School of Education.
- Wigelsworth, M., Humphrey, N., Kalamouka, A., & Lendrum, A. (2010). A review of key issues in the measurement of children's social and emotional skills. *Educational Psychology in Practice*, 26(2), 173-186.
- Zembar, M.J., & Blume, L.B. (2009). Gender and academic achievement. In Zembar, M.J., & Blume, L.B., *Middle childhood development: A conceptual approach (212-215)*. New York: Pearson
- Zinsser, K. M., Weissberg, R. P., & Dusenbury, L. (2013). *Aligning preschool through high school social and emotional learning standards: A critical and doable next step*. Chicago: Collaborative for Academic, Social, and Emotional Learning.

Table 1.

Definitions for the CASEL 5 Core SEL Competencies

Definitions	
Self-Awareness	The ability to accurately recognize one’s emotions and thoughts and their influence on behavior. This includes accurately assessing one’s strengths and limitations and possessing a well-ground sense of confidence and optimism.
Self-Management	The ability to regulate one’s emotions, thoughts, and behaviors effectively in different situations. This includes managing stress, controlling impulses, motivating oneself, and setting and working toward achieving personal and academic goals.
Social-Awareness	The ability to take the perspective of and empathize with others from diverse backgrounds and cultures, to understand social and ethical norms for behavior, and to recognize family, school, and community resources and supports.
Relationship Skills	The ability to establish and maintain healthy and rewarding relationships with diverse individuals and groups. This includes communicating clearly, listening actively, cooperating, resisting inappropriate social pressure, negotiating conflict constructively, and seeking and offering help when needed.
Responsible Decision-Making	The ability to make constructive and respectful choices about personal behavior and social interactions based on consideration of ethical standards, safety concerns, social norms, the realistic evaluation of consequences of various actions, and the well-being of self and others.

Table 2.

Student Demographic Descriptive Statistics (N =113)

	<i>N</i>	Percentage
Gender		
Male	53	47
Female	60	53
Grade		
Fourth	113	100
Ethnicity		
White/Non-Hispanic	55	49
Black/Non-Hispanic	24	21
Hispanic	14	12
American Indian/Alaskan	20	18
Native/Pacific Islander		

Table 3.

Typical Report Card Comment List of 24 Comment Options

Positive Comments	Negative Comments
1. Brings materials to class daily	13. Needs to bring materials to class regularly
2. Follows directions/rules well	14. Needs to follow directions/rules better
3. Presents work neatly	15. Needs to present work more neatly
4. Completes homework regularly	16. Needs to complete homework regularly
5. Completes class work regularly	17. Needs to complete class work regularly
6. Participates well in class discussions	18. Needs to participate in class discussions/group activities
7. Works well in a group	19. Needs to improve scores on tests
8. Works well independently	20. Needs to work more independently
9. Demonstrates continuous effort	21. Needs to pay closer attention in class
10. Demonstrates a positive attitude	22. Needs to demonstrate continuous effort
11. Shows improved behavior	23. Needs to complete projects on time
12. Conference requested by teacher	24. Needs to show more respect for others.

Table 4.

“Personal and Social Development” Rating Items

Likert Rating Items: N/A, check minus, check, check plus

1. Is able to follow classroom directions
 2. Is able to follow rules
 3. Is able to respect rights, feelings and property of others
 4. Is able to identify and accept responsibility for actions
 5. Is able to use B.E.S.T.* when addressing adults and peers
 6. Is able to use “Keep Calm”
 7. Is able to use Listening Position
 8. Is able to use Speaker Power
-

Table 5.

Data Analytic Techniques for the Current Study

	Data Analytic Techniques
Face & Content Validity	Qualitative Categorization Exploratory Factor Analysis
Reliability	Cronbach's Alpha
Demographic Bias	Independent Samples <i>t-tests</i>
Concurrent Validity	Hierarchical Multiple Regression
Utility	Descriptive Statistics

Table 6.

Conceptual Categorization of the 24 Typical Comments

	Self-Awareness & Self-Management	Social-Awareness & Relationship Skills	Non-SEL: Other
Positive	2. Follows directions/rules well 8. Works well independently 9. Demonstrates continuous effort 10. Demonstrates a positive attitude	6. Participates well in class discussions 7. Works well in a group 1. Brings materials to class daily 4. Completes homework regularly 5. Completes class work regularly	3. Presents work neatly 11. Shows improved behavior
Negative	14. Needs to follow directions/rules better 20. Needs to work more independently 21. Needs to pay closer attention in class 22. Needs to demonstrate continuous effort	18. Needs to participate in class discussion/group activities 24. Needs to show more respect for others 13. Needs to bring materials to class regularly 16. Needs to complete homework regularly 17. Needs to complete class work regularly 23. Needs to complete projects on time	12. Conference requested by teacher 15. Needs to present work more neatly 19. Needs to improve scores on tests

Table 7.

Conceptual Categorization of the “Personal and Social Development” Items

Self-Awareness & Self-Management	Social-Awareness & Relationship Skills	Responsible Decision Making
Follow classroom directions and rules	Respects rights, feelings, and property of others	Identifies and accepts responsibility for actions
Uses “Keep Calm”	Uses “B.E.S.T.” when addressing adults and peers	
Uses “Listening Position”	Uses “Speaker Power”	

Table 8.

*Descriptive Statistics for Cumulative Use of Comments across Marking Periods and Subject**Areas*

Comment	M	SD	Range
1	4.05	4.82	16
2	4.34	4.91	16
3	4.52	4.87	16
4	5.35	4.82	16
5	5.79	4.94	16
6	5.79	5.05	16
7	5.22	4.89	16
8	5.43	5.21	16
9	7.43	4.61	16
10	8.23	4.94	16
11	.35	.78	4
12	0	.00	0
13	.02	.13	1
14	.16	.56	3
15	.62	1.35	8
16	.87	1.68	11
17	.58	1.62	10
18	.11	.34	2
19	.71	1.22	6
20	.45	1.31	7
21	.72	1.67	10
22	.63	1.82	14
23	.12	.38	2
24	.03	.16	1

Table 9.

Exploratory Factor Analysis Component Loadings for Typical Comment Total Scores from All Marking Periods and Subject Areas

Comment	Component 1
1	.97
2	.97
3	.97
4	.95
5	.94
6	.93
7	.96
8	.96
9	.85
10	.79

Table 10.

Exploratory Factor Analysis Component Loadings for “Personal and Social Development”

Total Scores from All Marking Periods

Comment	Component 1
1	.89
2	.88
3	.92
4	.84
5	.86
6	.84
7	.84

Table 11.

Correlation Matrix for Variables Considered in Data Reduction Strategy

Variables	RW-Cum-GPA	SM-Cum-GPA	NJASK-LA-2007	NJASK-MA-2007	SaSM-Comm-Cum	SaRs-Comm-Cum	Pos-Comm-Total	SaSm-PSD-Cum	SaRs-PSD-Cum	Rdm-PSD-Cum	PSD-Total-Cum
RW-Cum-GPA	1.00										
SM-Cum-GPA	.86*	1.00									
NJASK-LA-2007	.62*	.55*	1.00								
NJASK-MA-2007	.67*	.72*	.68*	1.00							
SaSm-Comm-Cum	.62*	.57*	.42*	.45*	1.00						
SaRs-Comm-Cum	.57*	.50*	.37*	.39*	.95*	1.00					
Pos-Comm-Total	.60*	.52*	.40*	.42*	.98*	.99*	1.00				
SaSm-PSD-Cum	.49*	.46*	.37*	.33*	.37*	.28*	.32*	1.00			
SaRs-PSD-Cum	.45*	.42*	.42*	.37*	.37*	.30*	.33*	.88*	1.00		
Rdm-PSD-Cum	.43*	.37*	.41*	.36*	.31*	.22*	.26*	.84*	.87*	1.00	
PSD-Total-Cum	.48*	.45*	.41*	.36*	.38*	.29*	.33*	.97*	.97*	.91*	1.00

Note: RW-CumGPA = Cumulative GPA for Reading and Writing; SM-CumGPA = Cumulative GPA for Science and Math; NJAK-LA2007 = Scores on the NJASK for Language Arts; NJASK-MA2007 = Scores on the NJASK for Math; SaSMCommCum = Cumulative Comments related to Self-Awareness and Management; SaRsCummCum = Cumulative Comments related to Social Awareness and Relationship Skills; PosCommentTotal = Cumulative for comments one through ten out of the typical twenty-four comment options ; SaSmPSDCum = Cumulative for “Personal and Social Development” items related to Self-Awareness and Management; SaRsPSDCum = Cumulative for “Personal and Social Development” items related to Social-Awareness and Relationship; RdmPSDCum = Cumulative for “Personal and Social Development” items related to Responsible Decision-Making; PSDTotalCum = Cumulative for “Personal and Social Development” items

Correlational Ranges: 0.10-0.29 = Small, .30-0.49 = Medium, 0.50-69= Large, 0.70-0.89 = Very Large, 0.90-0.99 = Nearly Perfect

*Correlation is significant at the 0.05 level (2-tailed)

Table 12.

Comparison of Achievement Means for Male and Female Groups

	Male <i>(n = 53)</i> <i>M (SD)</i>	Female <i>(n = 60)</i> <i>M (SD)</i>	<i>p</i>
Cumulative GPA	3.44 (.39)	3.47 (.42)	.72
RWCumGPA	3.37 (.43)	3.44 (.44)	.42
SMCumGPA	3.51 (.39)	3.50 (.43)	.88
NJASK LA	219.35 (21.91)	223.54 (17.83)	.28
Scaled Score Range	Proficient	Proficient	-
NJASK MA	243.15 (31.21)	236.15 (32.26)	.26
Scaled Score Range	Proficient	Proficient	-

*Significant at the .05 level

**Significant at the .01 level

Note: NJASK ranges: 100-199 (Partially Proficient), 200-249 (Proficient), 250-300 (Advanced Proficient)

Table 13.

Descriptive Statistics for Achievement and Comment Variables

	<i>M</i>	<i>SD</i>	Range	Minimum	Maximum
CumGPA	3.45	.41	1.87	2.13	4.00
RW-CumGPA	3.40	.44	1.83	2.17	4.00
RW-GPAMP1	3.28	.52	2.33	1.67	4.00
RW-GPAMP4	3.54	.43	1.67	2.34	4.00
SM-CumGPA	3.50	.40	1.92	2.09	4.00
SM-GPAMP1	3.42	.46	1.92	2.09	4.00
SM-GPAMP4	3.49	.47	2.33	1.67	4.00
NJASK-LA	220.61	21.03	126	135	261
Scaled Score	Proficient	-	-	-	-
Range					
NJASK-MA	238.59	32.37	131	149	280
Scaled Score	Proficient	-	-	-	-
Range					
PosComment- Total	56.45	45.60	160	0	160
PSD-Total	93.10	12.61	50	62	112
PSD-Proficiency	11.32	10.31	28	.00	28
PSD-Deficiency	1.88	3.89	22	.00	22

Note: CumGPA=Cumulative GPA for All Subjects and Marking Periods

RW-CumGPA = Cumulative GPA for Reading and Writing

RW-GPAMP1 = GPA for Reading and Writing for Marking Period 1

RW-GPAMP4 = GPA for Reading and Writing for Marking Period 4

SM-CumGPA = Cumulative GPA for Science and Math

SM-GPAMP1 = GPA for Science and Math for Marking Period 1

SM-GPAMP4 = GPA for Science and Math for Marking Period 4

NJASK-LA2007 = Scores on the NJASK for Language Arts

NJASK-MA2007 = Scores on the NJASK for Math

PosCommentTotal = Cumulative for comments one through ten out of the typical twenty-four comment options;

PSDTotalCum = Cumulative for “Personal and Social Development” items

PSDProficiencyTotal = Cumulative score based on frequency of skills rated as check plus for the seven “Personal and Social Development Items across all marking periods

PSDDeficiencyTotal = Cumulative score based on frequency of skills rated as check minus for the seven “Personal and Social Development Items across all marking periods

Table 14.

Hierarchical Regression with Cumulative GPA as the Dependent Variable

	Standardized β		
	Model 1	Model 2	Model 3
Gender	.04	-.05	-.091
PosCommentTotal		.58***	.49***
PSDTotalCum			.33***
R^2	.01	.34***	.43***
R^2 Change	.00	.34***	.10***
F for change in R^2	.13	51.68***	17.20***

Note: PosCommentTotal = Cumulative score for comments assigned across all marking periods and subject areas
 PSDTotalCum = Cumulative score for the seven “Personal and Social Development Items across all marking periods

*Significant at the .05 level

**Significant at the .01 level

***Significant at the .001 level

Table 15.

Hierarchical Regression with Cumulative GPA for Reading and Writing as the Dependent

Variable

	Standardized β		
	Model 1	Model 2	Model 3
Gender	.08	-.00	-.05
PosCommentTotal		.59***	.50***
PSDTotalCum			.32***
R^2	.06	.35***	.44***
R^2 Change	.01	.34***	.09***
F for change in R^2	.13	53.57***	16.59***

Note: PosCommentTotal = Cumulative score for comments assigned across all marking periods and subject areas
 PSDTotalCum = Cumulative score for the seven "Personal and Social Development Items across all marking periods

*Significant at the .05 level

**Significant at the .01 level

***Significant at the .001 level

Table 16.

Hierarchical Regression with Cumulative GPA for Science and Math as the Dependent Variable

	Standardized β		
	Model 1	Model 2	Model 3
Gender	-.02	-.09	-.13
PosCommentTotal		.53***	.44***
PSDTotCum			.31***
R^2	.00	.28***	.37***
R^2 Change	.01	.28***	.09***
F for change in R^2	.03	39.86***	13.93***

Note: PosCommentTotal = Cumulative score for comments assigned across all marking periods and subject areas
 PSDTotCum = Cumulative score for the seven “Personal and Social Development Items across all marking periods

*Significant at the .05 level

**Significant at the .01 level

***Significant at the .001 level

Table 17.

Hierarchical Regression with NJASK Language Arts Scores as the Dependent Variable

	Standardized β		
	Model 1	Model 2	Model 3
Gender	.11	.05	-.12
PosCommentTotal		.38***	.29**
PSDTotalCum			.30**
R^2	.01	.15***	.23***
R^2 Change	.01	.14***	.08**
F for change in R^2	1.16	16.97***	10.54**

Note: PosCommentTotal = Cumulative score for comments assigned across all marking periods and subject areas
 PSDTotalCum = Cumulative score for the seven “Personal and Social Development Items across all marking periods

*Significant at the .05 level

**Significant at the .01 level

***Significant at the .001 level

Table 18.

Hierarchical Regression with NJASK Math Scores as the Dependent Variable

	Standardized β		
	Model 1	Model 2	Model 3
Gender	-.11	-.17	-.21
PosCommentTotal		.44***	.36***
PSDTotalCum			.29***
R^2	.01	.20***	.28***
R^2 Change	.01	.19***	.08***
F for change in R^2	1.29	24.50***	10.66***

Note: PosCommentTotal = Cumulative score for comments assigned across all marking periods and subject areas
 PSDTotalCum = Cumulative score for the seven “Personal and Social Development Items across all marking periods

*Significant at the .05 level

**Significant at the .01 level

***Significant at the .001 level

Table 19.

Hierarchical Regression with Cumulative GPA for Reading and Writing as the Dependent

Variable and PSD Proficiency then Deficiency Total Scores

	Standardized β			
	Model 1	Model 2	Model 3	Model 4
Gender	.07	-.01	-.05	-.09
PosCommentTotal		.59***	.54***	.48***
PSDProficiencyTotal			.27***	.20*
PSDDeficiencyTotal				-.23**
R^2	.01	.35***	.42***	.46***
R^2 Change	.01	.35***	.07***	.04**
F for change in R^2	1.29	24.50***	10.57***	

Note: PosCommentTotal = Cumulative score for comments assigned across all marking periods and subject areas
 PSDProficiencyTotal = Cumulative score based on frequency of skills rated as check plus for the seven “Personal

and Social Development Items across all marking periods

PSDDeficiencyTotal = Cumulative score based on frequency of skills rated as check minus for the seven “Personal and Social Development Items across all marking periods

PSDTotalCum = Cumulative score for the seven “Personal and Social Development Items across all marking periods

*Significant at the .05 level

**Significant at the .01 level

***Significant at the .001 level

Table 20.

Hierarchical Regression with Cumulative GPA for Science and Math as the Dependent Variable and PSD Proficiency then Deficiency Total Scores

	Standardized β			
	Model 1	Model 2	Model 3	Model 4
Gender	-.02	-.01	-.13	-.18
PosCommentTotal		.54***	.49***	.41***
PSDProficiencyTotal			.24**	.16*
PSDDeficiencyTotal				-.29**
R^2	.00	.28***	.34***	.40***
R^2 Change	.00	.28***	.05**	.06**
F for change in R^2	.61	40.13***	8.28***	10.67

Note: PosCommentTotal = Cumulative score for comments assigned across all marking periods and subject areas
 PSDProficiencyTotal = Cumulative score based on frequency of skills rated as check plus for the seven “Personal and Social Development Items across all marking periods
 PSDDeficiencyTotal = Cumulative score based on frequency of skills rated as check minus for the seven “Personal and Social Development Items across all marking periods
 PSDTotalCum = Cumulative score for the seven “Personal and Social Development Items across all marking periods
 *Significant at the .05 level
 **Significant at the .01 level
 ***Significant at the .001 level

Table 21.

Hierarchical Regression with Cumulative GPA for Reading and Writing as the Dependent

Variable and PSD Deficiency then Proficiency Total Scores

	Standardized β			
	Model 1	Model 2	Model 3	Model 4
Gender	.07	-.01	-.07	-.09
PosCommentTotal		.59***	.50***	.48***
PSDDeficiencyTotal			-.30***	-.23**
PSDProficiencyTotal				.20*
R^2	-.01	.35***	.42***	.46***
R^2 Change	.01	.35***	.07***	.03*
F for change in R^2	.49	54.30***	12.75***	6.34*

Note: PosCommentTotal = Cumulative score for comments assigned across all marking periods and subject areas
 PSDProficiencyTotal = Cumulative score based on frequency of skills rated as check plus for the seven “Personal and Social Development Items across all marking periods

PSDDeficiencyTotal = Cumulative score based on frequency of skills rated as check minus for the seven “Personal and Social Development Items across all marking periods

PSDTotalCum = Cumulative score for the seven “Personal and Social Development Items across all marking periods

*Significant at the .05 level

**Significant at the .01 level

***Significant at the .001 level

Table 22.

Hierarchical Regression with Cumulative GPA for Science and Math as the Dependent Variable and PSD Deficiency then Proficiency Total Scores

	Standardized β			
	Model 1	Model 2	Model 3	Model 4
Gender	-.02	-.10	-.17	-.18
PosCommentTotal		.54***	.49***	.41***
PSDDeficiencyTotal			-.34***	-.29**
PSDProficiencyTotal				.16
PSDTotalCum				
R^2	.00	.28***	.38***	.40***
R^2 Change	.00	.28***	.10***	.02
F for change in R^2	.49	40.129***	15.87***	3.47

Note: PosCommentTotal = Cumulative score for comments assigned across all marking periods and subject areas
 PSDProficiencyTotal = Cumulative score based on frequency of skills rated as check plus for the seven “Personal and Social Development Items across all marking periods
 PSDDeficiencyTotal = Cumulative score based on frequency of skills rated as check minus for the seven “Personal and Social Development Items across all marking periods
 PSDTotalCum = Cumulative score for the seven “Personal and Social Development Items across all marking periods
 *Significant at the .05 level
 **Significant at the .01 level
 ***Significant at the .001 level

Table 23.

Hierarchical Regression with NJASK Language Arts Scores as the Dependent Variable and PSD

Proficiency then Deficiency Total Scores

	Standardized β			
	Model 1	Model 2	Model 3	Model 4
Gender	.12	.06	.03	.00
PosCommentTotal		.38***	.33***	.28**
PSDProficiencyTotal			.25**	.21*
PSDDeficiencyTotal				-.16
R^2	.01	.15***	.21***	.23***
R^2 Change	.01	.14***	.06**	.02
F for change in R^2	1.42	16.96***	7.54**	2.55

Note: PosCommentTotal = Cumulative score for comments assigned across all marking periods and subject areas
 PSDProficiencyTotal = Cumulative score based on frequency of skills rated as check plus for the seven “Personal and Social Development Items across all marking periods

PSDDeficiencyTotal = Cumulative score based on frequency of skills rated as check minus for the seven “Personal and Social Development Items across all marking periods

PSDTotalCum = Cumulative score for the seven “Personal and Social Development Items across all marking periods

*Significant at the .05 level

**Significant at the .01 level

***Significant at the .001 level

Table 24.

Hierarchical Regression with NJASK Math Scores as the Dependent Variable and PSD

Proficiency then Deficiency Total Scores

	Standardized β			
	Model 1	Model 2	Model 3	Model 4
Gender	-.12	-.18	-.22	-.26
PosCommentTotal		.44***	.33***	.33***
PSDProficiencyTotal			.23**	.16
PSDDeficiencyTotal				-.25*
R^2	.02	.21***	.26***	.31***
R^2 Change	.02	.19***	.05**	.05*
F for change in R^2	1.55	24.80***	6.91**	6.72*

Note: PosCommentTotal = Cumulative score for comments assigned across all marking periods and subject areas
 PSDProficiencyTotal = Cumulative score based on frequency of skills rated as check plus for the seven “Personal and Social Development Items across all marking periods

PSDDeficiencyTotal = Cumulative score based on frequency of skills rated as check minus for the seven “Personal and Social Development Items across all marking periods

PSDTotalCum = Cumulative score for the seven “Personal and Social Development Items across all marking periods

*Significant at the .05 level

**Significant at the .01 level

***Significant at the .001 level

Table 25.

Hierarchical Regression with NJASK Language Arts Scores as the Dependent Variable and PSD

Deficiency then Proficiency Total Scores

	Standardized β			
	Model 1	Model 2	Model 3	Model 4
Gender	.12	.06	.02	.00
PosCommentTotal		.38***	.31**	.28**
PSDDeficiencyTotal			-.23*	-.16
PSDProficiencyTotal				.21*
R^2	.01	.15***	.20***	.23***
R^2 Change	.01	.14***	.04*	.04*
F for change in R^2	1.42	16.96***	5.42***	4.59

Note: PosCommentTotal = Cumulative score for comments assigned across all marking periods and subject areas
 PSDProficiencyTotal = Cumulative score based on frequency of skills rated as check plus for the seven “Personal and Social Development Items across all marking periods

PSDDeficiencyTotal = Cumulative score based on frequency of skills rated as check minus for the seven “Personal and Social Development Items across all marking periods

PSDTotalCum = Cumulative score for the seven “Personal and Social Development Items across all marking periods

*Significant at the .05 level

**Significant at the .01 level

***Significant at the .001 level

Table 26.

Hierarchical Regression with NJASK Math Scores as the Dependent Variable and PSD

Deficiency then Proficiency Total Scores

	Standardized β			
	Model 1	Model 2	Model 3	Model 4
Gender	-.12	-.18	-.22	-.26
PosCommentTotal		.44***	.35***	.33***
PSDDeficiencyTotal			-.30**	-.25*
PSDProficiencyTotal				.16
R^2	.02	.21***	.28***	.31***
R^2 Change	.01	.19***	.08**	.02
F for change in R^2	1.55	24.80***	10.65***	3.15

Note: PosCommentTotal = Cumulative score for comments assigned across all marking periods and subject areas
 PSDProficiencyTotal = Cumulative score based on frequency of skills rated as check plus for the seven “Personal and Social Development Items across all marking periods

PSDDeficiencyTotal = Cumulative score based on frequency of skills rated as check minus for the seven “Personal and Social Development Items across all marking periods

PSDTotalCum = Cumulative score for the seven “Personal and Social Development Items across all marking periods

*Significant at the .05 level

**Significant at the .01 level

***Significant at the .001 level

Table 27.

Hierarchical Regression with Combined GPA for Reading and Writing in Marking Period Four as the Dependent Variable while Controlling for GPA from Marking Period One

	Standardized β			
	Model 1	Model 2	Model 3	Model 4
Gender	.07	.00	.00	-.02
RW-GPAMP1		.73***	.73***	.67***
PosCommentTotal			.01	-.01
PSDTTotalCum				.18*
R^2	.01	.54***	.54***	.57***
R^2 Change	.01	.53***	.00	.03*
F for change in R^2	.61	40.129***	15.87	3.47*

Note: RW-GPAMP1 = Cumulative GPA for both Reading and Writing for marking period one

PosCommentTotal = Cumulative score for comments assigned across all marking periods and subject areas

PSDTTotalCum = Cumulative score for the seven “Personal and Social Development Items across all marking periods

*Significant at the .05 level

**Significant at the .01 level

***Significant at the .001 level

Table 28.

Hierarchical Regression with Combined GPA for Science and Math for Marking Period Four as the Dependent Variable while Controlling for GPA from Marking Period One

	Standardized β			
	Model 1	Model 2	Model 3	Model 4
Gender	-.04	-.08	-.11	-.13
SM-GPAMP1		.75***	.66***	.61***
PosCommentTotal			.22**	.19**
PSDTotalCum				.19**
R^2	.00	.56***	.58***	.62***
R^2 Change	.01	.55***	.04**	.03**
F for change in R^2	.20	128.48***	10.01**	7.54**

Note: SM-GPAMP1 = Cumulative GPA for Science and Math for marking period one

PosCommentTotal = Cumulative score for comments assigned across all marking periods and subject areas

PSDTotalCum = Cumulative score for the seven "Personal and Social Development Items across all marking periods

*Significant at the .05 level

**Significant at the .01 level

***Significant at the .001 level

Appendix A

CASEL 5: Domains, Definitions, and Examples

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

- 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
- 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 and relationship skills: 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; 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.

- 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
- being polite
- bystander awareness
- 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)

Appendix B

De-identified Copy of a Fourth Grade Report Card

MARKING PERIOD	1	2	3	4
LANGUAGE ARTS / LITERACY				
Reading	A	A	A	A
Applies a variety of reading strategies to construct meaning (predicts, uses illustrations, makes connections)	✓	✓	✓	✓
Comprehends what is read	✓	✓	✓	✓
Learns and uses new words	✓	✓	✓	✓
Adjusts reading strategies based on purpose (enjoyment, learning, problem solving)	✓	✓	✓	✓
Uses a variety of sources to gather information and draw conclusions	✓	✓	✓	✓
Comments: 5, 6, 9 / 1-10 / 1-10 / 1-10				
Writing	A	A	A	A
Uses the writing process (draft/revise/edit)	✓	✓	✓	✓
Understands the differences in writing for a variety of purposes and audiences (inform, entertain, respond to literature, poetry)	✓	✓	✓	✓
Uses correct punctuation and capitalization	✓	✓	✓	✓
Applies spelling skills	✓	✓	✓	✓
Writes legibly using proper size and spacing	✓	✓	✓	✓
Comments: 8, 9, 10 / 1-10 / 1-10 / 1-10				
Listening / Speaking				
Expresses ideas clearly	✓	✓	✓	✓
Listens and responds actively	✓	✓	✓	✓
Participates in group discussions	✓	✓	✓	✓
Comments:				
SCIENCE				
Understands topics and concepts	✓	✓	✓	✓
Ability to raise questions about the world and willingness to seek answers	✓	✓	✓	✓
Applies process skills (observing, classifying, predicting and questioning)	✓	✓	✓	✓
Accurately interprets and analyzes data in activities/experiments	✓	✓	✓	✓
Works cooperatively in completing experiments	✓	✓	✓	✓
Comments: 1-10 / 1-10 / 1-10 / 1-10				

MARKING PERIOD	1	2	3	4
MATHEMATICS				
Understands topics and concepts	A	A	A	A
Applies operations accurately (computation and estimation)	✓	✓	✓	✓
Is able to collect, organize, and interpret data	✓	✓	✓	✓
Selects and uses a variety of appropriate problem solving strategies (pictures, words, numbers, reasonableness of answers, and diagrams)	✓	✓	✓	✓
Communicates mathematical thinking clearly both in written and oral form	✓	✓	✓	✓
Demonstrates proficiency with basic facts	✓	✓	✓	✓
Comments: 3, 4, 7, 10, 14, 21 / 6, 9, 10, 14, 11 / 1-11 / 1-8, 11, 21, 10				
SOCIAL STUDIES				
Understands topics and concepts (civics, economics, history)	A	A	A	A
Compare and contrast (culture, time periods, events, causes, effects)	✓	✓	✓	✓
Understands the cause and effect relationship of events	✓	✓	✓	✓
Gathers and summarizes useful information from a variety of sources	✓	✓	✓	✓
Applies and uses geographic skills and concepts (maps, globes, etc.)	✓	✓	✓	✓
Comments: 1-10 / 1-10 / 1-10 / 1-10				
PERSONAL / SOCIAL DEVELOPMENT				
Follows classroom directions and rules	✓	✓	✓	✓
Respects rights, feelings and property of others	✓	✓	✓	✓
Identifies and accepts responsibility for actions	✓	✓	✓	✓
Uses B.E.S.T. when addressing adults and peers	✓	✓	✓	✓
Uses "Keep Calm"	✓	✓	✓	✓
Uses Listening Position	✓	✓	✓	✓
Uses Speaker Power	✓	✓	✓	✓

Performance Scale

A Outstanding
 B Very good
 C Satisfactory
 N Not yet meeting expectations

✓+ Consistently exceeds expectations
 ✓ Consistently meets expectations
 ✓- Does not consistently meet expectations
 NA Not applicable at this time

Comments

1 Brings material to class daily	13 Needs to bring materials to class regularly
2 Follows directions / rules well	14 Needs to follow directions / rules better
3 Presents work neatly	15 Needs to present work more neatly
4 Completes homework regularly	16 Needs to complete homework regularly
5 Completes class work regularly	17 Needs to complete class work regularly
6 Participates well in class discussions	18 Needs to participate in class discussions / group activities
7 Works well in a group	19 Needs to improve scores on tests
8 Works well independently	20 Needs to work more independently
9 Demonstrates continuous effort	21 Needs to pay closer attention in class
10 Demonstrates a positive attitude	22 Needs to demonstrate continuous effort
11 Shows improved behavior	23 Needs to complete projects on time
12 Conference requested by teacher	24 Needs to show more respect for others

Attendance	1	2	3	4	Total
Days Absent	2	0	0	0	2 / 182
Times Tardy	0	0	0	0	0 / 182
Absenteeism/Tardiness has affected progress					