Children at risk of academic failure: How child health and social-emotional skills affect reading and mathematics achievement from kindergarten through fifth grade

by

Julia Parkinson

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Education)
in The University of Michigan
2011

Doctoral Committee:

Professor Brian P. Rowan, Chair
Professor Joanne F. Carlisle
Professor David L. Featherman
Professor Bendek B. Hansen

brave reader
if you can make it through
I dedicate this dissertation
wholeheartedly
to you

Acknowledgements

Thank you to my advisor, Brian Rowan, and to the other members of my dissertation committee – David Featherman, Ben Hansen, and Joanne Carlisle. Thanks to Robin Jacob for being a great mentor. And finally, a big thank you to my family and friends for their love and support.

Table of Contents

Dedication ... ii

Acknowledgements ... iii

List of Tables ... vii

List of Figures ... viii

Abstract ... ix

Chapter

1. Introduction ... 1

Statement of Problem ... 1

Description of the Dissertation ... 4

Description of Data and Data Analysis Procedures ... 6

Limitations ... 10

Organization of the Dissertation ... 13

2. Models of Child Development ... 14

Conceptual Model ... 15

Ecological Systems ... 16

Development over Time ... 17

3. Literature Review ... 19

Definitions ... 19

Physical Health ... 20

Mental Health and Social-emotional Skills ... 21

Academic Achievement ... 28

Relationship of Health and Social-emotional Skills to Academic Achievement ... 28

Effect of Physical Health on Academic Achievement ... 29

Effect of Social-emotional and Regulation Skills on Academic Achievement 30

Summary 36

4. Methodology 37

Data Source 37

Study Sample 38

Missing Data 38

Measures 45

Measures of Child Academic Achievement 45

Measures of Child Social-emotional Skills....47

Measures of Child Health 60

Measures of the Child, Family, and Home 61

Measures of the Teacher, Classroom, and School 62

Data Analysis Procedures 63

HCM of Math and Reading Achievement 63

Effect Size Computation 68

Testing the Sensitivity of the HCM Reading and Math Models to Specification Error 69

Modeling Proficiency Scores 71

5. Results 73

Overview 73

Research Question 1: Effects of Health and Social-Emotional Skills on Reading and Math Achievement over Time 75

Reading Achievement Model Results....75

Math Achievement Model Results....85

Sensitivity of Math and Reading Results to Specification Error 94

Research Question 2: Exploring the Combined Effect of Multiple Risks on Reading and Math Achievement....96

6. Discussion and Conclusions 103

Social-emotional Skills 105

Stability of Measures over Time....106

Effect of Regulatory Behaviors on Reading and Math Achievement....108

Effect of Internalizing Problems on Reading and Math Achievement 111

Physical Health 112

Gaps by Risk Group 113

Directions for Future Research 117

Appendix 119

Bibliography 139

List of Tables

Table

1.1 Measures of child health and social-emotional skills 4

4.1 Description of eligible and ineligible children in ECLS-K 39

4.2 Mover attrition over time in ECLS-K..40

4.3 Unstandardized mean differences and standard errors for eligible and ineligible ECLS-K children in fifth grade 41

4.4 Completion rates over time in ECLS-K...42

4.5 Stability of Social Rating Scale factors over time 57

4.6 Method for accumulating repeated child measures over time when creating a time-varying covariate...59

4.7 Growth rate values used to calculate achievement status at each time point...............65

4.8 Specification of model coefficients 67

4.9 Reading and math proficiency levels..72

5.1 Unstandardized coefficients and standard errors for covariates of interest from HCM models predicting kindergarten to fifth grade reading achievement................77

5.2 Effect sizes over time of health and social-emotional skills measures and other key child characteristics on reading achievement 81

5.3 Unstandardized coefficients and standard errors for covariates of interest from HCM models predicting kindergarten to fifth grade mathematics achievement........86

5.4 Effect sizes over time of health and social-emotional skills measures and other key child characteristics on math achievement 90

5.5 Means and standard deviations of kindergarten child, family, teacher and school risk factors by regulatory behaviors risk group 98

5.6 Reading achievement gaps between regulatory behaviors risk groups 100

5.7 Math achievement gaps between regulatory behaviors risk groups 101

List of Figures

Figure

1.1 Diagram of methods to model the effects of repeated measures over time on academic achievement........8

1.2 Total effect size of regulatory behaviors on reading achievement from four separate HLM models........9

2.1 Conceptual model of factors leading to academic achievement........16

4.1 Total effect size of regulatory behaviors on reading achievement from four separate HLM models........59

5.1 The standardized effect size of health and social-emotional skills measures on reading achievement from kindergarten through fifth grade........82

5.2 Predicted cumulative probability of fifth grade reading proficiency by average regulatory behaviors........83

5.3 The standardized effect size of health and social-emotional skills measures on math achievement from kindergarten through fifth grade........91

5.4 Predicted cumulative probability of fifth grade math proficiency by average regulatory behaviors........92

5.5 Effect of average regulatory behaviors on reading achievement from two statistical models with different dependent measures of achievement and a third model using propensity score strata........95

5.6 Effect of average regulatory behaviors on reading achievement from two statistical models with different dependent measures of achievement and a third model using propensity score strata........95

5.7 Average reading achievement trajectories by regulatory behaviors risk group........100

5.8 Average math achievement trajectories by regulatory behaviors risk group........102

Abstract

One of the most enduring problems facing educators in America is that a large number of children underachieve academically relative to their peers upon school entry and continue to underachieve throughout their schooling. It is widely assumed that knowing what child factors cause academic success or failure can eventually lead to the development of interventions that target those child factors and improve the academic achievement of children at risk. This dissertation explores how children's health and social-emotional skills affect their reading and math achievement trajectories over the elementary school years. Specifically it investigates which constructs within these two domains have the strongest effect on reading and math achievement, and the size of these effect over time, when accounting for repeated measures of health and social-emotional skills.

Hierarchical cross-classified models (HCM) of reading and math achievement from kindergarten through fifth grade are created to explore these issues, using data from the Early Childhood Longitudinal Study-Kindergarten Class of 1998-1999 (ECLS-K). The cross-classified structure of these models accounts for the different schools ECLS-K children moved through over the elementary school years. Repeated measures of health and social-emotional skills are included in the HCM models either as time-varying covariates or as stable traits.

The study finds that many of the child health and social-emotional skills measures included have statistically significant effects on reading and math achievement at kindergarten entry and throughout elementary school, holding all other measures in the model constant. The measure of children's regulatory behaviors (a factor based on the average of four highly correlated scales from the teacher social rating scale) has the strongest independent effect on achievement of all health and social-emotional skills measures, as well as of most of the background control measures included in the model. The standardized effect sizes of the regulatory behaviors factor on reading and math

achievement at kindergarten entry is 0.18 and 0.23, respectively, and increases to 0.26 and 0.29, respectively, by fifth grade. The regulatory behaviors factor accounts for almost half of the total achievement gap (calculated using all covariates in the model) between children with poor regulatory behaviors and the average child.

Chapter 1

Introduction

Statement of Problem

One of the most enduring problems facing educators in America is that a large number of children underachieve academically relative to their peers upon school entry and continue to underachieve throughout their schooling. This finding begs the question of how educators can improve the academic performance of underachieving children (see, for example, Burchinal, Peisner-Feinberg, Pianta, & Howes, 2002; Gutman, Sameroff, & Cole, 2003; Jencks & Phillips, 1998; Rothstein, 2004; Shonkoff & Phillips, 2000). Despite decades of research and a host of variably successful academic interventions, a large proportion of students remain at risk of poor academic achievement. For example, data from the 2009 National Assessment of Educational Progress (NAEP) show that 33% of fourth graders read below the basic fourth-grade reading level and 18% have below basic grade-level math ability (NCES, 2009).

The risk of poor academic achievement is even greater for children who come from socially and economically disadvantaged backgrounds or from racial or ethnic minority groups. The NAEP 2009 fourth-grade data also shows that 22% of White students versus 52% of Black students scored below basic reading levels. Using the Early Childhood Longitudinal Study–Kindergarten Cohort (ECLS-K) data, I investigated these raw achievement gaps for different groups of students over time (USDE NCES, 2006). For example, Black students entered kindergarten 0.4 standard deviations below their White peers on the ECLS-K reading achievement test, but by the end of fifth grade, they were 0.8 sd behind.[1] This corresponds to Black students being, on average, academically

[1] This gap was calculated using mean differences on achievement scores from the raw data. Model-based estimates of the Black/White gap tend to be smaller after controlling for other measures for the child, family, and school.

one school-year behind their White peers by fifth grade. Similarly large gaps exist between children from economically disadvantaged backgrounds and their advantaged peers.

The size of the achievement gaps just described and the large number of children affected by them have motivated large expenditures of resources among researchers and educators in an effort to reduce or eliminate them. Because no single factor causes achievement gaps and underachievement, efforts to overcome these problems must involve identifying the major contributing factors and designing interventions specifically targeted at those multiple factors. Decades of research have identified numerous factors affecting academic achievement including the child's own characteristics and skills (such as IQ, health, self-regulation, behavior, and cognitive skills), the family, home and neighborhood (such as parental investments, cognitive stimulation in the home, and food insecurity), and the classroom and school (such as school resources, teacher-child relationship, and teacher quality) (Borman, Hewes, Overman, & Brown, 2002; Brooks-Gunn & Duncan, 1997; Brooks-Gunn, Duncan, & Aber, 1997; Brooks-Gunn & Markman, 2005; Cook & Evans, 2000; Currie, 2005; Dauber & Epstein, 1993; Evans, 2004; Feuerstein, 2000; Foster, 2002; Garrett, Ng'andu, & Ferron, 1994; Gutman et al., 2003; Hedges, Laine & Greenwald, 1994; Henderson & Berla, 1994; Jackson, 2003; Jencks & Phillips, 1998; Klebanov, Brooks- Gunn, McCarton, & McCormick, 1998; McLeod & Kaiser, 2004; McLoyd, 1998; Raudenbush, 2004; Rivkin, Hanushek, & Kain, 2005; Rothstein, 2004; Rowan, Correnti, & Miller, 2002; Scott-Jones, 1995; Spira & Fischel, 2005; Trzesniewski, Moffitt, Caspi, Taylor, & Maughan, 2006).

In this dissertation I focus on how measures of child health and social-emotional skills affect the academic achievement of elementary school children, while taking into account other aspects of the child, home, and school. Child health, both physical and mental, and child social-emotional skills have gained increased attention from both researchers and educators over the past few decades. In addition to being important developmental outcomes in their own right, many researchers have argued that good health and social-emotional competence are as important for academic success as are academic/cognitive skills (Heckman, 2006; Raver, 2004; Raver & Zigler, 1997; Shonkoff & Phillips, 2000). In fact, researchers have consistently found that health and social-emotional skills are correlated with various measures of academic achievement (Arnold, 1997; Currie, 2005; Evans, 2004; Gutman et al., 2003; Kagan, Moore, & Bredekamp, 1995; McLeod & Kaiser, 2004; Rothstein, 2004; Spira & Fischel, 2005; Trzesniewski et al., 2006).

While the importance of health and social-emotional skills and their association with academic achievement are generally accepted, there is still a lot that is unknown. First, there has been little consensus on what specific aspects of child health and social-emotional skills are associated with academic achievement. The difficulty in reaching a consensus is due in part to the multi-dimensional nature of the domains being studied. Both health and social-emotional skills encompass a large number of constructs with overlapping definitions and measures that come from a wide variety of fields (e.g. Blair & Diamond, 2008; Graziano, Reavis, Kaene & Calkins, 2007; McLeod & Fettes, 2007; Shields et al., 2001). Most researchers ignore the multi-dimensionality of these domains by focusing on only one type of measure or measures of one construct at a time. This approach has lead to inaccurate findings on both the size and significance of effects (see Duncan et al., 2007 as an example of contradictory findings when including multiple covariates in the same model). In an effort address the multi-dimensionality of health and social-emotional skills and to determine which of the two has the greatest effect on achievement, I include in this dissertation multiple measures from both the domains.

Another shortcoming in the published studies is the lack of longitudinal studies on the effect of child health and social-emotional skills on academic achievement. Historically, most researchers that model academic achievement have used measures of health and social-emotional skills from a single point in time, primarily from preschool and kindergarten entry. Developmental models of health and social-emotional skills, however, strongly suggest that these areas are not constant during middle childhood. Accurately understanding their effect on achievement requires accounting for the variability of these constructs over time. In this dissertation I include repeated measures of health and social-emotional skills in order to account for the time-varying nature of these constructs.

Description of the Dissertation

In this dissertation I explore how child health and social-emotional skills affect reading and mathematics achievement trajectories over the elementary school years. In exploring this relationship I make two contributions to the existing literature. First, I provide a broad picture of the domains of health and social-emotional skills and their association with academic achievement by reviewing the literature in various fields, primarily psychology and education. Second, I include my own secondary analysis of data to explore the longitudinal relationship between various measures of health and social-emotional skills and math and reading achievement.

My literature review elaborates on the various constructs that have been discussed within the domains of health and social-emotional skills and how definitions and constructs overlap with each other. I also briefly review how constructs within these domains are measured and the difficulties in analysis and interpretation of results due to measurement issues.

My longitudinal analysis draws on data from the Early Childhood Longitudinal Study-Kindergarten (ECLS-K), and is therefore limited to measurements from that study, which are listed in Table 1.1. While some of these measures are fairly broad, covering an array of skills (such as the regulatory behaviors factor), they can still be useful in pointing researchers to areas that deserve more precise focus in future research. While a number of indicators of academic success or failure are commonly recognized, such as grades or graduation, I use reading and math achievement as an indicator of what children have

Table 1.1
Measures of child health and social-emotional skills

Health Measures	Social-Emotional Skills Measures
Overall Health Scale	Internalizing Problems
Disability Status	Regulatory Behaviors (factor combining externalizing problems, self-control, interpersonal skills, and approaches to learning scales)
Birth Weight	
Premature	
Body-Mass Index (BMI)	
Food Insecurity	
Doctor/Dentist Visit in Past Year	
Health Insurance	

learned. The measures of academic achievement in the ECLS-K come from standardized achievement tests and can be scaled over time, making them ideal for longitudinal models.

I address the following questions in my secondary data analyses:

1) How do health and social-emotional skills affect reading and math achievement over time, controlling for other child, family, home and school factors?
2) What specific factors are associated with low social-emotional skill levels, and how large are the effects of these factors when combined with other factors on children's academic achievement over time?

With this first question I am interested in first determining how health and social-emotional skills affect reading and math achievement when all other measures are controlled for in the model. Second, I am interested in the size of these effects from kindergarten through fifth grade. Specifically, I want to know whether a significant effect at kindergarten entry exists for health and social-emotional skills, and whether, over time, the size of the effect on reading and math increases, decreases or stays about the same. Additionally, I want to explore how the measurements of the effects of health and social-emotional skills compare to the measurements of other control measures commonly cited in the literature.

With the second research question, I took a step back from looking at the effects of individual measures on achievement in order to gain a sense of how all risk factors, in combination, affect a child's achievement. First, I am interested in determining what other risk factors children who are already at risk for poor self-regulation might have. Second, through creating profiles of groups of children with different overall risk levels, I can determine the size of the total reading and math achievement gaps between risk groups from kindergarten through fifth grade. I can then determine for each risk group what proportion of the total achievement gap is due to child health and social-emotional skills measures. Such an analysis will provide a more accurate picture of how important these domains are to academic achievement.

Description of Data and Data Analyses Procedures

To answer both of these research questions, I used data from the Early Childhood Longitudinal Study–Kindergarten Class of 1998–1999 (ECLS–K) (USDE NCES, 2006), a study sponsored by the National Center for Education Statistics of the U.S. Department of Education. The ECLS–K is a large, ongoing, longitudinal study of the academic development of roughly 22,000 children who passed from kindergarten to fifth grade in U.S. schools between the years 1998 and 2003 (Rathbun & West, 2004). The ECLS–K data allowed me to analyze academic achievement of elementary school-age children over time and to compare these data with variables such as levels of health and social emotional functioning, family and social backgrounds, and quality of schools.

I analyzed ECLS-K data using a hierarchical, cross-classified longitudinal model (HCM) (Raudenbush & Bryk 2002, pg. 373). The ECLS-K sample included children who stayed at the same school from kindergarten through fifth grade and children who moved to different schools during that period. I deviated from standard hierarchical linear modeling of children nested in schools over time, because such a technique would limit the sample to children who remained in the same school at each time period modeled. Excluding movers from the sample would drop close to half of the sample and could add significant bias to the model estimates. Conversely, the HCM approach involves cross-classifying children within schools, thus allowing me to retain both movers and non-movers in the dataset.

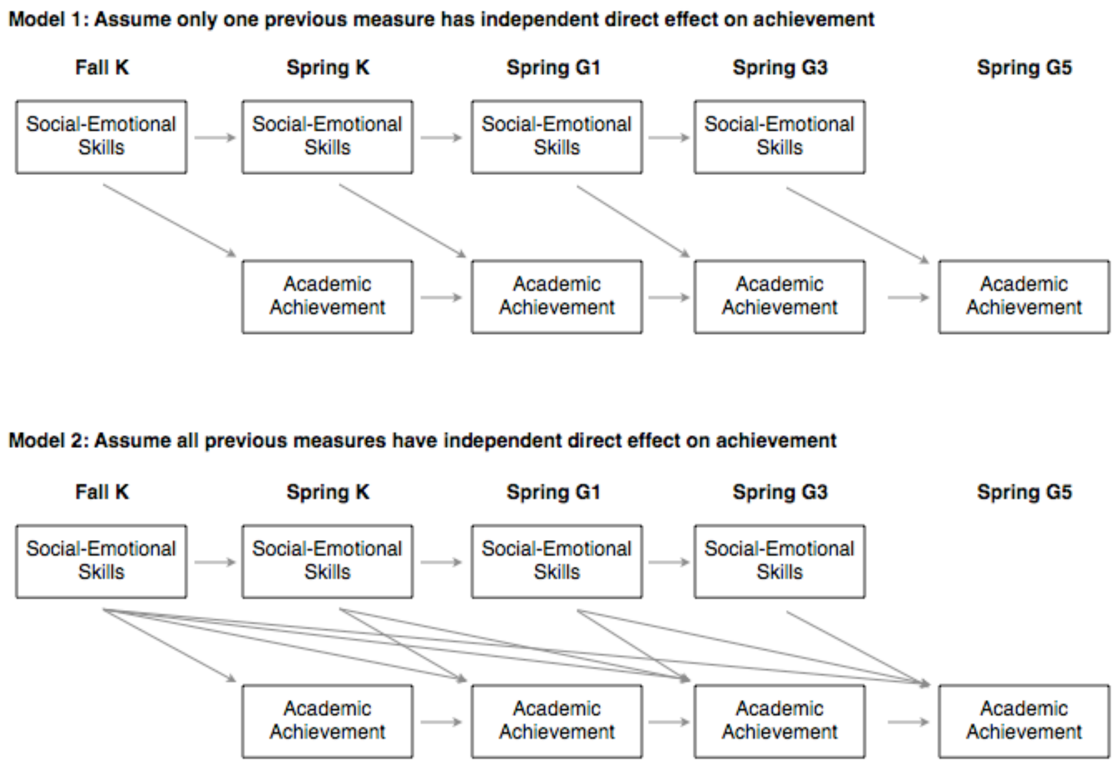
I modeled math and reading achievement from kindergarten through fifth grade using four separate growth rates, corresponding to the data collection periods of the ECLS-K. These include: growth during kindergarten, growth from the end of kindergarten to the end of first grade, growth from the end of first grade to the end of third grade, and growth from the end of third grade to the end of fifth grade. Using separate growth rates for these time periods provided model estimates of academic achievement over time that closely corresponded to actual achievement at each time point in the data.

My analyses of math and reading achievement allowed me to address a number of weaknesses in the current literature. First, rather than focusing on a single construct at a time, I included a wide range of health and social-emotional skills measures in my achievement models, as listed in Table 1.1. I found the measures within each domain to be correlated with each other and with achievement. Excluding some of these measures from my achievement models could potentially lead to inaccurate estimates of the coefficient size and significance of the measures included in the model. Including all of these measures thus allowed me to determine, in a broad sense, which measure had the strongest effect on achievement and the size of that effect. To reduce the threat of confounding variables, I also included in my model a large number of individual and contextual covariates as controls.

Most longitudinal models of academic achievement use a single time point (typically preschool or kindergarten entry) to assess of health and social-emotional skills (e.g. Claessens, Duncan & Engel, 2008; Raver, Garner, & Smith-Donald, 2007). Because both of these domains fluctuate over time, as described in Chapter 3, this approach can lead to underestimations of the true effect of health and social-emotional skills on academic achievement. My approach is to include measurements of health and social-emotional skills over multiple time points. Two possible methods to such an approach exist. The first method would be to include the most recent measure (time t-1) as a predictor of achievement at each time point (see Figure 1.1, model 1). This method assumes that a given measures of social-emotional skills has a direct effect only on the period directly following that measure.

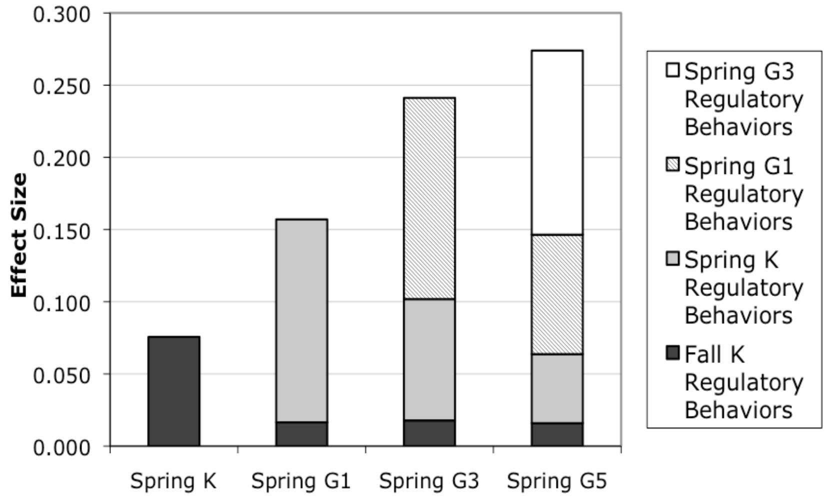
Alternatively, the second method assumes that prior measures of social-emotional skills have a direct long-term effect on achievement. Under this method, all prior measures of social-emotional skills are used to model academic achievement at a given time point (Figure 1.1, model 2) I chose to use the second method of modeling in my analysis because prior research has shown that effects from a single point in time can have a lasting significant effect on outcomes for at least 2-3 years (Barnett, 1995; Currie & Stabile, 2006; Gutman et al., 2003; Palloni, 2006). For example, Claessens and colleagues (2008) found that kindergarten measures of attention from the ECLS-K still had a statistically significant effect on reading and math achievement by fifth grade.

Figure 1.1
Diagram of methods to model the effects of repeated measures over time on academic achievement



In addition to the findings in the literature, my own simple analysis using ECLS-K data shows that early measures of health and social-emotional skills maintained a long-term effect on achievement even after accounting for later measures in the model. I did my analysis by looking at a series of 2-level (children nested within schools) hierarchical linear models. The first model had kindergarten reading achievement as the dependent variable, and the other three models had first, third, and fifth grade reading achievement as their outcome, respectively. In each of these models, I included as predictors all measures taken of health and social-emotional skills measures prior to the grade of the outcome, as well as a large set of control variables. I found that even after controlling for the most recently measured regulation factor, all but the kindergarten entry measures remained significant in each model. Figure 1.2 shows a graphical depiction of the effect size of all the measures of regulatory behaviors in these four models. I found that the effects of the earlier measures did eventually diminish, but remained significant for from one to five years, even after including more recent measures in the model. This evidence

Figure 1.2
Total effect size of regulatory behaviors on reading achievement from four separate HLM models



validated my measures in the model. This evidence validated my choice to include all prior measures in my longitudinal model since doing so would lead to more accurate estimates of the size of the effects of social-emotional skills and health measures on achievement. I would also expect to find larger effect sizes from those reported in previous research that has included only early measures from a single point in time.

In order to include the information from multiple measures over time of a single construct in my reading and math longitudinal models, I needed to create a single time-varying covariate for each construct. I could then include this time-varying covariate in the level 1 (achievement scores over time) portion of the HCM models and, as the name suggests, the value of the covariate would vary over time. In order to account for the effect of all prior measures at a give time point, t, I summed the measures from kindergarten entry to time t-1. Using this accumulated measure was similar to including each of the prior measures independently, as in the HLM models described above. I did not include concurrent measures of health and social-emotional skills in order to maintain their temporal precedence to measures of academic achievement. Measures I included as time-varying covariates were interpersonal skills, regulatory behaviors, overall health, disability status, visits to doctor or dentist and health insurance.

Limitations

Using the ECLS-K data for my analyses provided me with sufficient data to increase the power of finding results and also allowed for the exploration of effects on very different children from a diversity of backgrounds. However, there were a number of limitations to using this data. One limitation was that I was constrained by the type and quality of measures available in the ECLS-K. While the ECLS-K had a large range of health measures and ratings of behavior on 5 scales, these measures were general in scope (e.g., parent ratings of "overall health status"). This lack of measurement precision limits my ability to interpret or make claims about the results. This concern is most relevant for the regulatory behaviors factor. In creating this factor I based my assumptions on a credible statistical analysis and evidence from prior research; however, if these assumptions are inaccurate, my regulatory behaviors factor may reflect an unknown underlying construct. The creation of the regulatory behaviors factor and rationale for my assumption that this factor reflects children's regulation is described in chapter 4.

Related to the problem of measurement precision, is the issue of correctly identifying the construct underlying what is being rated by the parents and teachers. Part of the difficulty lies in the inconsistent use of terminology in the literature to describe both the constructs and the measures. For example, in the ECLS-K, the teachers answer a series of questions regarding the child's outward behaviors which have been compiled into a scale termed externalizing behaviors. This term, however, has not only been used in the literature to describe similar ratings of behavior by observers, but more recently for direct measures of children's externalizing behaviors. While there is evidence that these measures are correlated, the measures are not necessarily indicative of the same underlying construct, and interpretations and extrapolations should be made with caution. Interpretations of the results from my analyses, and comparing these results to those of other studies, must not be done solely based on the names of the measures, but based on how the measures were collected.

Even with these limitations, results from my analyses using these measures are useful, particularly in comparison to the existing literature. General measures, such as those used in the ECLS-K, are common in large, multi-purpose, survey datasets. Despite their broad nature, these general measures provide an estimate of the importance of the area being measured, thus providing a starting point for researchers to investigate using more precise measures. Because of the limitations that I have explained, the results and conclusions presented in this dissertation, particularly those based on the regulation factor, should be verified using datasets with more precise, direct measures of the underlying constructs.

Another limitation of much of the ECLS-K data is the large measurement error. Measurement error is the difference between a child's true score (such as their true level of overall health) and the child's estimated score (parent rating of overall health). Measurement error can arise from both random and systematic sources. One of the major reasons for systematic error in these measures is parent and teacher bias. Ratings of children by any observer in any dataset, are prone to bias due to the observers' motivations, beliefs, feelings towards the child, memory, and comparisons with other children (De Los Reyes & Kazdin, 2005; Yougstrom, Loeber, & Stouthamer-Loeber, 2000). Although I was unable to determine precisely how much of the variability in the ECLS-K measures was due to error, a simple measurement model for my self-regulation factor showed that nearly 38% of the variability in the factor was from error. I expect that most of measures that I use from the ECLS-K data (e.g., regulation, internalizing problems, and overall health) to also have substantial amount of error. Measurement error could lead to bias in my model estimates and could also mask the true effects of these measures on achievement.

A final concern with my analyses was my limited ability to make causal claims. Ideally, analyses like this one would determining what characteristics of the child and what aspects of the environment cause developmental outcomes; however, in practice, determining causality is very difficult. Determining whether a certain variable (such as regulation) has a causal effect on an outcome (such as reading achievement) requires knowing two things. First, one must know the child's reading achievement based on their given regulation. Second, one must know the counterfactual, or the potential outcome the child would have received if they had a different level of regulation (Holland, 1986). This counterfactual, unfortunately, is impossible to observe. Determining causal effects requires comparing groups of children that are similar on both observed and unobserved measures. Randomized experiments allow for such comparisons since one could make the assumption that unmeasured variables are equal; however, most educational research is based on observational data, allowing researches to control for only observed characteristics. Unobserved characteristics still pose a serious threat to any causal claims. Statistical methods have been developed for use with observational data in order to address the problems of unobserved covariates. Methods such as matching, propensity scores, and instrumental variables allow researchers to come closer to inferring causation (Rosenbaum, 2002).

In this dissertation I use a number of strategies to strengthen causal claims of the effects of health and social-emotional skills on academic achievement. First, my model only includes predictors that have temporal precedence on the outcomes (Raudenbush, 2001). Second, because I was not using randomized experimental data, I tried to account for all potentially confounding variables by including multiple covariates in my model, many of which have been found to be correlated with achievement. Another method I used to limit the effects of confounding variables was propensity score stratification. A propensity score measures the propensity a child has for receiving a given treatment. Propensity scores are calculated by predicting treatment status from a wide range of measured covariates. A study sample can be divided, based on the propensity scores, into balanced strata where each of the strata contains children from both the treatment and control groups with similar measured characteristics, and hopefully similar unmeasured characteristics. Dummy variables for these propensity strata can then be included in achievement models. While there was no actual treatment in this study, I modeled the propensity for children having self-regulation then ran an additional reading and math HCM model including propensity strata based on the regulation propensity scores.

By maintaining the temporal precedence of predictors, including a large array of covariates in my model, and re-testing the model with propensity strata, I have decreased the likelihood of major specification errors in my results. I can therefore claim weak causality of health and social-emotional skills on reading and math achievement.

Organization of the Dissertation

I describe my dissertation research beginning in Chapter 2, which is a brief overview of child development theories and my own conceptual model that guided my secondary data analyses. In Chapter 3, I review the literature on health and social-emotional skills, first providing definitions, and then turning to the relationship of these two domains to children's academic achievement. In Chapter 4, I summarize the data, measures, and analytic methods I employed in my secondary analyses of ECLS-K data. I present the results of these analyses in Chapter 5. In Chapter 6, I discuss these findings and their implications, elaborating on the conclusions I can draw from my work.

Chapter 2

Models of Child Development

The first few years of life are a period of dramatic development. Development is described as an accumulation of experiences in which children are active participants (Shonkoff & Phillips, 2000). Through infancy and early childhood, children begin to understand and regulate emotions, they begin to develop relationships with peers and learn basic social skills, and they also begin to develop basic language, literacy, and communication skills. This development continues as children enter middle childhood and transition into school.

Research has shown that developmental problems experienced during early and middle childhood can set in motion patterns that can have detrimental effects into adulthood (Moffit, Caspi, Harrington, & Milne, 2002, Palloni, 2006; Sameroff, 1986, Strohschein, 2005). There is a general consensus in various academic disciplines (including psychology, sociology, economics and education) that child development is a complex process involving the child and their family, neighborhood, school, community and beyond (Bronfenbrenner, 1977 & 1979; Bronfenbrenner & Morris, 1998; Heckman, 2006; Sameroff & Fiese, 2000, Schulenberg, Maggs & O'Malley, 2003).

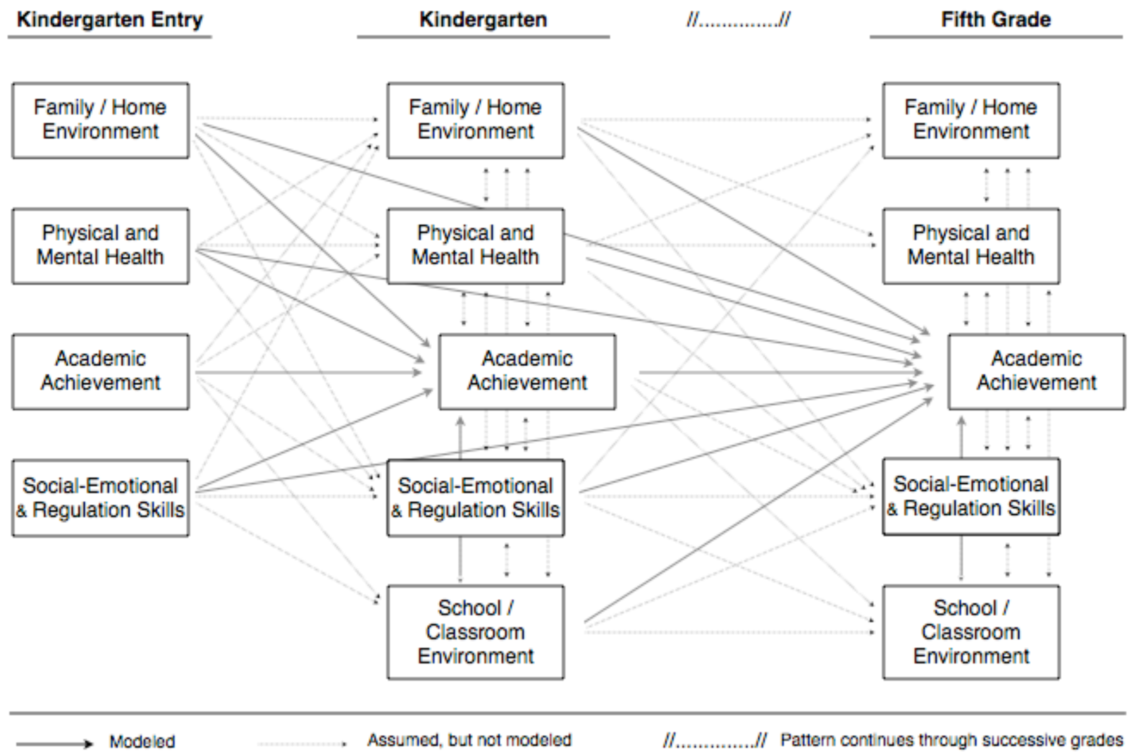
According to ecological and transactional models of child development, a given domain of development, such as academic development, does not occur independently, but must be seen as part of a larger developmental system (Essex et al., 2006; Sameroff, 2000). Each child's academic development is a complex function of the child's own cognition, language, social and emotional competencies, regulation, and health, as well as the surrounding ecological system comprised most directly of parents, family, community, and schools.

Conceptual Model

In this dissertation I focus on two child domains that have been shown to affect academic development in elementary school: (a) health (both physical and mental); and (b) social-emotional skills (e.g. Currie, 2005; Hinshaw, 1992b; Miles & Stipeck, 2006; Morrison, Ponitz & McClelland, 2009; Raver, Garner, & Smith-Donald, 2007). Figure 2.1 provides a diagram of how health and social-emotional skills, as well as other areas of the ecological system, affect academic achievement. As shown in this figure, I begin tracking the relationship of achievement with child health and social-emotional skills in kindergarten. While health and social-emotional skills are associated with academic development prior to kindergarten, the start of formal schooling marks a key transitional period and is a critical time to begin exploring these relationships in more detail. The skills children have at school entry not only provide a starting point for further development, but also influence the initial interactions children have with teachers, and can therefore affect the teachers' perceptions of that child throughout the school year (Baker, 2006; Baker, Grant & Morlock, 2008). A survey of kindergarten teachers found that they consider social and behavioral skills (such as working independently and in groups, resolving conflicts, and following directions) to be equally important for success in school as cognitive and academic skills (such as communication, language, and literacy) (Rimm-Kaufman, Pianta & Cox, 2000).

In my conceptual model illustrated in Figure 2.1, I present a variety of interactions between factors that influence a child's academic achievement. I assume that both prior and current health, social-emotional skills, and other child, family, and school factors influence achievement. However, in order to maintain temporal precedence, I do not model all of these possible pathways in my analytic model. The dark solid lines in Figure 2.1 indicate pathways I include in my analytic model, while the grey dotted lines are pathways I assume are occurring but did not test. The vertical lines in a given year indicate transactional processes that could be occurring throughout that year within the child's own skills or between the child and his/her home and school environments. Transactions are reciprocal interactions of children with their environment and the people in the environment (Sameroff, 2000). Transactions occur when the child's skill (such as behavior) changes aspects of the environment (such as teacher warmth towards the child),

Figure 2.1
Conceptual model of factors leading to academic achievement



which in turn affects the child's skill (Sameroff & Mackenie, 2003). While I believe these bi-directional processes diagramed in my conceptual model are occurring, I do not test these transactions in my analytic model due to the complexity required to create a model that could test all possible pathways of influence.

In developing this conceptual model, I took into account ecological theories of development as well as theories of skill development over time. I outline these theories briefly below, describing the concepts from these theories that I incorporated into my conceptual model.

Ecological Systems

Ecological theories of development outline the nested contexts surrounding children (Bronfenbrenner, 1977). The primary context within which the child develops is comprised of the home and school environments (Bronfenbrenner, 1979; Bronfenbrenner, 1977; Sameroff & Fiese, 2000). Historically, family and home environments have been seen as the major source of inequality in children's academic achievement, beginning with the publication of the Coleman report in 1966. Using techniques such as value-added modeling, education researchers have found that schools and teachers are also a major source of variation in academic achievement, together accounting for up to 30-60% of the variability in reading and math achievement (Konstantopoulos, 2006; Nye, Konstantopoulos, & Hedges, 2004; Rowan et al., 2002).

Although the home and school are the primary environments within which children develop and are most likely to have a direct effect on academic achievement, a number of other parts of the ecosystem also have both direct and indirect effects. These include the broader environments of neighborhoods and governments, as well as less-obvious factors such as societal norms, ideology, and laws. Accurately modeling academic achievement, therefore, requires consideration of not only the child, but also the home and school environments, as well as the more distant components of the ecosystem.

In my conceptual model I included a number of elements of the larger ecosystem, including the child's family background, home environment, and school experiences that are also important factors affecting academic achievement. Although I did not closely investigate the relationships of these factors with achievement for this dissertation, I did control for them in my model. It is important to note that my model includes only the child and context measures that were available in the ECLS-K. To represent the home context, I included measures of material hardship, parental investments, parent stress, parent expectation, home environment, and positive parenting. The classroom context is represented by measures of classroom composition, overall classroom behavior, instructional practices, teacher characteristics, and parent participation at school. Last, the school context is represented by measures of school structural features, school composition, and school policies. Because of data limitations, I included only a few distal measures of the ecosystem; these measures include neighborhood safety, school policies and the geographical region of the school.

Development over Time

An obvious, but often ignored, component of developmental models is the dynamic nature of the child's developmental context and the changes in the child's own skills and assets, such as social skills, behavior problems or health status (Pianta, 2006). The skill levels, interactions and experiences children have over time, can cause developmental trajectories to become reinforced, mediated or reversed (Heckman, 2006; Schulenberg et al., 2003; Shonkoff & Phillips, 2000; Sroufe, 1996). In their model of skill formation, Cunha & Heckman (2007) outline mechanisms through which early skills and experiences can affect the development of later skills. They argue that skills at stage t+1 are a function of all past investments, added through time t (Cunha & Heckman, 2007 & 2008; Heckman, 2007). Investments can include the child's own skills as well as investments from the child's surrounding environments (i.e. family, home, school). One could add to this model other characteristics and processes of the child, family, and school that might not be considered investments but have an impact on achievement.

Cunha and Heckman asserted that skills produced at an early stage enhance skills learned at later stages, implying that skills are self-reinforcing, cross-fertilizing, and long-lasting (see also Duncan et al., 2007; Raver, Gershoff & Aber, 2007). This effect can be seen, for example, in early childhood literature. At-risk children begin school with significant gaps in academic achievement, and these gaps continue to grow over time (Duncan et al., 2007; Parkinson & Rowan, 2008). Intervening early to improve children's academic skills has been shown not only to improve children's school readiness skills, thus reducing the gap at school entry, but also to improve academic achievement over the first few years of elementary school (Barnett, 1995; Barnett, Hustedt, Robin, & Schulman, 2004; Cunha, Heckman, Lochner & Masterov, 2006; NICHD ECCRN 2004; Peisner-Feinberg et al., 2001).

In my conceptual model, I account for the changing nature of children's health and social-emotional skills over time, as well as the changing home and school context. As suggested by Cunha & Heckman (2007), I assumed academic achievement was a function of all past measures, beginning from kindergarten entry, of child health and social-emotional skills.

Chapter 3
Literature Review

The domains of health and social-emotional skills are broad and widely studied by researchers from a number of different fields. In this chapter I first attempt to align various definitions of constructs within the health and social-emotional skills domains. I then summarize the literature on the association of health and social-emotional skills with academic achievement.

Definitions

Health has been broadly defined as "the extent to which children are able or enabled to develop and realize their potential, satisfy their needs, and develop the capacities that allow them to interact successfully with the biological, physical, and social environment" (NRCIM, 2004, p. 4). As can be seen from this definition, health is a multidimensional concept. In practice, it is generally broken out into two general domains: physical health and mental health (Palloni, 2006; Stevens, 2006). Conceptualizations of mental health are often closely linked to constructs from another domain commonly studied by researchers—social-emotional skills. Children showing competence in social-emotional skills can be thought of as having mental health, while children with behavior problems have mental health problems. Regulation skills, in turn, could be the underlying foundation for all of these outward behaviors. In this section I attempt to consolidate how these terms are used and conceptualized in the literature. I begin with physical health, then jointly discuss mental health and social-emotional skills. Lastly, I turn to my outcome of interest—academic achievement.

Physical Health

Physical health refers to the health of the individual's physical body. The term most commonly used to represent physical health in the literature is health status. In children, this is typically measured by adult-report of the child's health based on a simple rating scale. Health status can also refer to other measures of health such as birth weight and the presence of a chronic condition (Currie, 2005; Palloni, 2006). In his investigations into the long-term effects of health, Palloni outlines the complexity of health status. He argues that a child's health status is formed early in life. Health status, he suggests, consists of factors beginning from in utero and extending to the levels and rate of change of physical growth and development, to exposure to and contraction of acute and chronic illnesses, and finally to general fitness, frailty, energy and alertness.

Actual measures of physical health, however, are significantly more limited. Researchers are typically restricted to measures such as birth weight, whether babies are born prematurely, malnutrition, obesity, general reports of health status, and the presence of chronic conditions. Chronic health conditions include problems such as asthma, poor hearing, dental carries (tooth decay), allergies, and ear infections (Case, Lubotsky & Paxson, 2002; Currie, 2005). These last three are the most common chronic conditions found among children, with higher rates found among children in poverty (Currie, 2005). Parents generally report high levels of overall health status. Data from the National Survey of Early Childhood Health show only 15% of parents reported children having good, fair, or poor health status (reported by Stevens, 2006).

There are a number of risk factors associated with poor health. Poverty is the most commonly cited and studied risk factor for health, with reviews consistently finding that increased poverty is associated with more health problems (for reviews, see Aber, Bennett, Conley, & Li, 1997; Bradley & Corwyn, 2002; Case et al., 2002; Goldman, 2001). Not only are wealthier children more likely to have better overall heath, they are less likely to develop chronic health conditions and have been shown to be less affected by and recover more quickly from early health problems (Case et al., 2002). Other risk factors for poor health which are correlated with poverty include environmental exposures; parent physical and mental health; home, neighborhood, and school safety; access to health care; food insecurity and malnutrition; lack of exercise and education

(Adler & Ostrove, 1999; Case & Paxson, 2006; Currie, 2005; Duncan & Brooks-Gun, 2000; Essex et al., 2006; Fiscella & Kitzman, 2009; Palloni, 2006; Shonkoff & Phillips, 2000; Stevens, 2006; Xue, Leventhal, Brooks-Gunn, & Earls, 2005). These risk factors often co-occur leading to higher odds of poor health and developmental delays, with at least one-third of American children having two or more risk factors for poor health (Stevens, 2006).

Mental Health and Social-emotional Skills

Mental health and social-emotional skills are often used in the literature as catchall phrases for labeling development skills encompassing social competence and problem behaviors to attention, self-regulation, and executive functioning skills. The large array of terms are sometimes defined clearly, but more often are used loosely and sometimes interchangeably to describe the underlying skills or observed behavior of children. In this section I first review how mental health is defined in the literature, followed by definitions of social-emotional skills and how these constructs map onto each other.

Mental Health

Mental health is an integral part of a child's development and well-being (US DHHS, 1999). It is defined as how people think, feel, and act in life's situations. Children with mental health are successful in achieving expected developmental, social, emotional and cognitive milestones, developing secure attachments and social relationships, and coping with adversity and stress (US DHHS 1999; Berlin, Brooks-Gunn, McCarton, & McCormick, 1998). One of the difficulties in understanding mental health, is that it requires a value judgment that varies not only across cultures but even across individuals within cultures, based on individual beliefs and perceptions (US DHHS 2001).

While mental health is an overall measure of well-being, children are typically identified as having or lacking mental health based on their outward behaviors. When behaviors deviate from what is expected, they are referred to as mental health problems. Once these problems become serious and reach a clinically diagnosable level, they are called mental disorders. Not all behavior problems, however, reach a diagnosable level, though they still reflect deviations from normal behavior (Xue et al., 2005). Mental health problems are signs and symptoms without the intensity or duration needed to meet the criteria for a mental disorder, but which still affect a person's thoughts, feelings and behaviors and can be thought of as occupying the middle ground between mental health and mental illness (US DHHS, 1999).

Mental health problems are typically classified into externalizing and internalizing behavior problem domains (Achenback, 1966, Merrell, 2003). Externalizing behavior occurs when children have a problem inhibiting socially prohibited behavior. Behaviors include aggression, antisocial behavior, defiance and oppositionality, hyperactivity, impulsivity, over-activity, and inattention (Essex et al., 2006, Hinshaw, 1992a, Xue et al., 2005). Internalizing behavior, on the other hand, occurs when children inhibit behavior too much, resulting in depression and sadness, anxiety, fearfulness, social withdrawal such as shyness and timidity, and other behavior inhibitions (Essex et al., 2006, Xue et al., 2005). Internalizing behaviors are harder to identify than externalizing behaviors, because there are fewer outwardly observable characteristics.

Social-emotional Skills

Educators tend to refer to mental health and mental health problems with terms associated with social-emotional development. Social-emotional development has been broadly defined as the ability to adjust internal processes, such as thoughts and emotions, to the demands of the surrounding environment (Blair, 2002). This development requires the production of social-emotional skills. Whether or not children have learned and developed these skills is known by observing how they function (behave) within specific contexts, as social-emotional skills are dependent on the context within which the child functions. A child's observed social functioning can vary widely across situations and settings, as a child learns how to use and apply his or her social skills in different contexts (Fischer, Bullock, Rotenberg & Raya, 1993). I use the term social-emotional skills to refer to the abilities and subsequent social behaviors of children.

I have grouped the skills representing social-emotional development into three distinct constructs, consistent with the literature: problem behaviors, social skills, and learning-related social skills (Ladd, Herald, & Kochel, 2006; McClelland & Morrison, 2003; Raver & Zigler, 1997). Problem behaviors are identical to the construct of mental health problems, and are classified as externalizing and internalizing behavior problems. Interpersonal skills, also referred to as social skills, social or prosocial behavior, and social functioning, are defined as the knowledge and ability to use social behaviors appropriate to various interpersonal situations, to navigate social relationships, and to interact well with others (Gresham & Elliot, 1990; Welsh & Bierman, 2001). Social skills represent a broad range of skills and behaviors, including interacting positively with peers, sharing, cooperating, listening and communicating, respecting and helping others, and community-building. The third construct, learning related social skills, reflects social behaviors in interpersonal settings that are closely related to learning, and stem from regulatory skills such as attention, memory, and inhibition (McClelland, Cameron, Wanless, & Murray, 2007; McClelland, Acock, & Morrison, 2006). Observable skills in this domain include the ability to listen, remember, and follow directions; communicate effectively; participate well in groups; and stay on task (McClelland, Acock, & Morrison, 2006).

Social competence is a term frequently used in connection with social skills and functioning. Competence has been broadly defined as the upper limit of a child's level of development, observed by the child's ability to do specific development tasks and to adapt and reach high levels of functioning in specific situations and environments (Fischer et al., 1993; Masten & Coatsworth, 1998; Welsh & Bierman, 2001). Competence can be achieved in social, emotional, cognitive and academic domains. Having social competence implies high levels of social functioning, or social effectiveness, such as having and maintaining positive relationships, following rules, and being socially responsible (Welsh & Bierman, 2001). This definition of social competence ties in closely with the earlier definition of mental health, which stated that mental health is the child's success at achieving expected developmental, social, and emotional milestones. Having social competence, then, is an indicator of also having mental health.

For the remainder of this chapter, I use the term social-emotional skills to refer to children's social, emotional and mental health skills. I use the term competence to refer to successful functioning in a given context, and behavior problems to refer to externalizing and internalizing problems.

Self-Regulation

Over the past decade, many researchers have argued that the skills required for social-emotional competence are founded on self-regulation skills, due to the strong correlation between the constructs and how they conceptually build on each other (Masten & Coatsworth, 1998; Rothbart & Posner, 2005; Rueda, Posner, & Rothbart, 2005; Wentzel, 1991). Shonkoff & Phillips (2000) maintain that self-regulation is a cornerstone of childhood development, cutting across and influencing all domains of behavior. A growing body of research has found support for this theory, finding that higher levels of self-regulation are associated with improved measures of social-emotional skills, from higher social skills and social competence to lower problem behaviors (Bronson, 2000; Cameron, Connor & Morrison, 2005; Colman, Hardy, Albert, Raffaneli, & Crocket, 2006; Eisenberg, Champion & Ma, 2004; Kochanska & Knaack, 2003; Masten & Coatsworth, 1998; Spinrad et al., 2006; Wentzel, 1991). While this association suggests that self-regulation may underlie social-emotional skills, the specific role components of regulation play and the direction of causal pathways between these areas is not fully understood or agreed on. I have chosen in this dissertation to agree with the argument by Shonkoff & Phillips (2000) that self-regulation is the foundation of social-emotional development.

As with definitions of social-emotional skills, self-regulation has not been defined or operationalized in a consistent manner in the literature (Eisenberg et al., 2004; Patrick, 1997). I have chosen to use the working definition that self-regulation is a collection of tools enabling individuals to exercise control over, direct, and plan their emotions, cognitions, behaviors, and attention (Morrison et al., 2009; Blair & Diamond 2008; Cole, Martin, & Dennis, 2004; Kochanska, Murray, & Harlan, 2000; Rueda et al., 2005). Regulation skills include the ability to shift and control attention, inhibit inappropriate responses while activating appropriate responses, and use memory. Different aspects of self-regulation that have received attention in the literature include emotion regulation and effortful control, which form the temperamental base of regulation (Cole et al., 2004; Eisenberg et al., 2004; Kochanska et al., 2000; Murray & Kochanska, 2002; Rothbart & Posner, 2005; Rothbart & Bates, 2006; Rueda et al., 2005; Spinrad et al., 2006); behavior regulation (Howse, Calkins, Anastopoulos, Keane, & Shelton, 2003; McClelland, Cameron, Connor et al., 2007; Ponitz, McClelland, Matthews, & Morrison, 2009); and executive processes, otherwise known as executive functions (Blair, 2002; Cole & Deater-Deckard, 2009; Diamond, 2006; Kerr & Zelazo, 2004; Zelazo & Muller, 2002). Executive function skills are the regulatory processes required to plan and execute goal-directed behavior, such as working memory, attention control and shifting, and inhibition. These skills have been described as the building blocks of self-regulation as well as of academic and cognitive skills (Blair & Razza, 2007; Gathercole & Baddeley, 1993).

Prevalence

Of the three areas of social-emotional skills, behavior problems has historically received the most attention. In fact, prevalence rates are generally only reported for children with problems, not for those demonstrating social competence. Behavior problems are reported to affect at least one in every five children, based on data from the Methodology for Epidemiology of Mental Disorders in Children and Adolescents (MECA) study, with 5% of parents reporting children with severe problems, such as Attention Deficit Hyperactivity Disorder (ADHD) (Pastor, Reuben, & Falkenstern, 2004, US DHHS, 2001). The most commonly reported behavior problems are difficulty sitting still, inability to take turns, and interrupting, with higher reports for children in poverty (Currie, 2005). Children from low-income families and boys have higher prevalence rates of behavior problems (Brooks-Gunn, Duncan, & Aber, 1997; Costello et al., 1996; Essex et al., 2006; Najman et al., 2004; Pastor et al., 2004; Yu & Williams, 1999), while reports of prevalence rates by race have been mixed (Costello et al., 1996; Pastor et al., 2004; Strohschein, 2005). Risk factors for behavior problems are similar to those for poor health, and include parent mental health, family stress, the child's own temperament, as well as teacher and classroom processes and school success (Essex et al., 2006).

Development

Literature on the development of social-emotional skills has focused largely on the development and change over time of behavior problems. Determining whether change has actually occurred, however, is challenging, due to the difficulty of determining whether the problem itself has changed or just the outward expression of the problem has changed (Achenbach, 2005). For example, results from some research studies show stable or increasing problems throughout childhood (Brame, Nagin & Tremblay, 2001; Campbell, Shaw & Gilliom, 2000; Henricsson & Rydell, 2006; Hinshaw, 1992b; Kowaleski-Jones & Duncan, 1999; Silver et al., 2005) while others maintain that rates decrease over time (Bub, McCartney & Willett, 2007; Keenan, Shaw, Delliquadri, Giovannelli, & Walsh, 1998; Keiley, Bates, Dodge, & Pettit, 2000; Tremblay, 2000). Other researchers have found that problem trajectories differ for different groups of students (McLeod & Fettes, 2007; Zhou et al., 2007). For example, a review of externalizing behavior problems found that problems deepen into disorders for some children and weaken in others and the longer a child displays symptoms in early childhood, the more likely these behaviors will intensify into disorders by later childhood (Spira & Fischel, 2005).

Other researchers have looked at change by examining the rank order of individuals over time. These researchers have found that while the surface manifestations of externalizing behavior problems are inconsistent, the rank order of individuals with problems is maintained over time (Hinshaw, 2002; Lahey, McBurnett, & Loeber, 2000; Patterson, Forgatch, Yoerger, & Stoolmiller, 1998). One of the reasons for this stability in rank order over time could be due to the consistent levels of risk in a child's encompassing environments (Caspi, 2000).

Context, in fact, may play a large role in the perceived stability or instability of social-emotional skills. Researchers have found that children's level of functioning changes as the context changes (Fischer et al., 1993; Hinshaw, 2002; Masten & Coatsworth, 1998). Children moving from one context to another, such as from home to school or from one grade to another, might demonstrate different levels of functioning and also might require time to adjust their skills to the new context. Elements of the context that might influence functioning include relationships with parents, teachers and peers; parent mental health and other home stressors; access to mental health services; and poverty with its associated hardships (Ackerman, Brown, & Izard, 2003; Alzabo-Poria, Pike & Deater-Deckard, 2004; Colman et al., 2006; Johnson, McGue, Iacono, 2006; Lahey et al., 1995; Loeber, Farrington, Stauthamer-Loebner, & Van Kammen, 1998; Raver, Gershoff, & Aber, 2007; Repetti, Taylor, Seeman, 2002; Strohschein, 2005; Vazsonyi & Huang, 2010).

Evidence from intervention research provides additional insight into the change of social-emotional skills over time. This research base has found that all three areas of social-emotional skills, as well as regulatory skills, are modifiable. Social-emotional learning (SEL) interventions, such as PATHS and 4Rs, have shown that social and emotional skills can be improved through a combination of instruction and practice (Domitrovich, Cortes, & Greenberg, 2007; Jones, Brown, Hoglund, & Aber, under review; Raver et al., 2009). Researchers at CASEL recently performed a meta-analysis of hundreds of SEL studies and found overall effect sizes on social emotional skills of .60, indicating significant growth in these skills (Payton et al., 2008). Though the evidence is still sparse, training regulatory skills, especially executive functioning skills, also leads to improved self-regulation, with the most effective interventions teaching and practicing these skills directly (Diamond, Barnett, Thomas, & Munro, 2007; Jacob & Parkinson, working paper).

Measurement

The multidimensionality of children's social-emotional skills makes it difficult to measure these skills in practice. Some measures attempt to capture overall social competence or problem behaviors, while others tap into more specific aspects of social-emotional skills. Researchers, however, often use different measures to represent similar constructs, though the measures might tap into different skills and domains. For example, social competence has been measured using a child's sociometric status, teacher/parent/self-ratings of socially appropriate behavior, and the number of friends the child has.

Over the past few decades, a wide variety of mental health screening and assessment tools have been developed for use by researchers in an attempt to reliably capture the various dimensions of mental health problems (see Levitt, Saka, Romanelli, & Hoagwood, 2007 for a review). These tools typically consist of sets of questions for parents, teachers or children, that are combined to form scales for specific behavior problems or for more general domains of problems. Because there are no definitive assessment procedures or clear thresholds for levels of severity, making assessments is highly subjective and challenging.

Other issues that must be considered when measuring and using measures of social-emotional skills are the age of the child, timing of the problem, observation context, and potential biases of the rater (De Los Reyes & Kazdin, 2005; Richters, 1992; Youngstrom et al., 2000). I examine some of these issues in more depth in Chapter 4.

Academic Achievement

My outcome of interest in this dissertation is academic achievement. Academic achievement is defined as the level of attainment achieved by students in a given academic subject, or as the level of proficiency in regards to specific standards of achievement. In other words, academic achievement indicates what children have learned or what they can do with what they have learned about academic subjects, such as reading, math or science. Another term commonly used by education researchers is academic growth. While academic achievement is a reflection of what the child knows, academic growth indicates the change in achievement levels over time. Growth does not take into account the initial or ending achievement levels of students, only the amount of change over time.

Academic achievement is typically measured using standardized achievement tests. Other measures of achievement include school grades and teacher surveys or ratings of what an individual child knows. Some researchers in the literature use more distal measures of general success or failure in schools as proxies to learning. These are attendance, retention, suspension, high school graduation or dropout, and college entrance.

Relationships of Health and Social-emotional Skills to Academic Achievement

Now that I have described what is typically meant by physical and mental health and social-emotional skills in the literature, I turn to how measures of these constructs affect the academic achievement of children in elementary school. Although this review focuses on school-aged children, it should be noted that measures of good health as well as social-emotional and regulatory competence are associated with overall positive development throughout the entire life-course (Shonkoff & Phillips, 2000; McLeod & Kaiser, 2004; Palloni, 2006; Roeser & Eccles, 1998). In fact, Palloni (2006) found that childhood physical and mental heath not only affected adult health but also adult social class accession, position, and earnings, at rates similar to more conventional measures (see also Cutler & Lleras-Muney, 2006).

The focus of this dissertation, however, is only on the effects of measures of health and social-emotional skills on achievement during elementary school, though I also include some evidence from preschool, middle school and high school studies. I begin this section by summarizing the research on the relationship between measures of physical health and academic achievement, followed by the evidence linking measures of social-emotional and regulation skills to academic achievement.

Effect of Physical Health on Academic Achievement

Research has found associations between multiple aspects of physical health and academic achievement. Measures of health from birth—birth weight, being born preterm, and having other natal difficulties—are consistently found to be associated with academic development. There is disagreement, however, on how long this association lasts, with some studies showing persistent negative effects through elementary school on cognitive development (such as IQ) and academic development, particularly in reading (Case, Fertig & Paxson, 2005; Casey, Whiteside-Mansell, Barrett, Bradley & Gargus, 2006; Caughy, 1996; Reichman, 2005). Other studies have found that early deficits due to problems at birth are overcome within the first few years (Case et al., 2002; Wilson, 1985).

General measures of early health, physical activity, and nutrition are also associated with reading and math achievement; while poor nutrition, lack of exercise, and poor health are associated with lower language, reading and math skills over the elementary school years (Glewwe, Jaboy & King, 2001; Hanson, Austin & Lee-Bayha, 2004; Powell, Walker, Chang, & Grantham-McGregor 1998). Some reasons proposed for the relationship between measures of health and academic achievement are that illnesses could alter body chemistry; illness could take children out of school and take time away from other educational activities; chronic conditions could lead to more stress, fatigue or pain, which could negatively influence academic and cognitive development; and illness could alter the relationship between children and their parents, peers, and teachers (Case et al., 2002; Currie, 2005; Currie & Stabile, 2003; Newacheck, Jameson, & Halfon, 1994).

Measures of chronic conditions, such as asthma, have also been shown to relate to academic achievement (for reviews, see Case & Paxson, 2006; Currie 2005). These chronic conditions are not always severe enough to limit activity, though they might affect behavior, energy levels, and attention, which in turn could affect achievement (Currie, 2005).

Effect of Social-emotional and Regulation Skills on Academic Achievement

The effect of measures of mental health and social-emotional and regulatory skills on achievement is generally believed to be more important than the effect of physical health (Bronson, 2000; Graziano et al., 2007; Hair, Halle, Terry-Humen, Lavelle, & Calkin, 2006; Howse, Calkins et al., 2003; Malecki & Elliot, 2002; Morrison et al., 2009; Raver, Garner, & Smith-Donald, 2007; Raver & Zigler, 1997). Different measures of skill and achievement, the covariates controlled for, and statistical methods used, however, have led to differences in the size, significance, and direction of this association. There is also an ongoing debate over which skills are most important. Before summarizing this literature, I briefly review another debate in the literature concerning the direction of causality between social-emotional skills and academic achievement.

Researchers are divided on how social skills and behavior problems are related to academic achievement (Hinshaw, 1992b; Miles & Stipek, 2006; Spira & Fischel 2005). Using longitudinal data, some researches have found evidence that prior behavior problems lead to later reading difficulties (Rapport, Denney, Chung & Hustace, 2001; Roeser, Eccles & Strobel, 1998, Spira & Fischel, 2005), while others have found that prior reading difficulties lead to later behavior problems (Arnold & Doctoroff, 2003; Bennett, Brown, Boyle, Racine, & Offord, 2003; Roeser et al., 1998; Silver, Measelle, Armstrong & Essex, 2005). Others contend that some underlying factor, or common cause, such as attention skills, explains the development of both outcomes, such as attention skills (Trzesniewski et al., 2006). In reality, I would expect that behavior and academic development are reciprocal processes, which interact with each other, with additional child skills, and with the context in which the child is developing (McGee, Prior, Williams, Smart, & Sanson, 2002; Rabiner, Coie, & CPPRG, 2000; Trzesniewski et al., 2006). My literature review focuses on how social skills and behavior problems affect academic achievement. I begin by reviewing studies of behavior problems, followed by studies of social skills, and then turn to studies looking at learning-related skills and self-regulation. Most of the studies in each of these areas are only able to report whether or not there was an association between these variables, but are not able to determine why that association existed or what the causal pathways really are. While many hypotheses exist on what these causal pathways are and why these associations exist, I focus my review on evidence concerning whether or not associations exist between measures of my constructs of interest and academic achievement.

Comprehensive reviews of the literature on how measures of behavior problems (mostly focusing on externalizing behavior problems) are associated with academic achievement have found that measures of inattention, hyperactivity, aggression, and antisocial behavior are all associated with lower levels of emergent literacy skills in preschool and lower levels of math and reading achievement throughout elementary, middle and high school (Arnold, 1997, Doctoroff, Greer & Arnold, 2006, Dobbs, Doctoroff, Fisher, & Arnold, 2006, Hinshaw, 1992b; Rowe & Rowe, 1999; Spira & Fischel, 2005). Long-term follow-up studies have found that children demonstrating externalizing problem behaviors at some point during their education (elementary, middle school or high school) were less likely to complete high school or enter college. It was unclear from the research, however, whether it was the duration of the problem, or if the child had a problem at any point in time that had a greater impact on academic outcomes such as achievement, retention, and drop-out rates. For example, one study found that children with persistent behavior problems were at greater risk of high school dropout than those who only had problems at one point in time (McLeod & Fettes, 2007). Flanagan and colleagues (2003), on the other hand, found that having high risk of problem behaviors at any time during elementary school was just as bad as having problem behaviors over all time points, based on teacher ratings of academic performance.

The relationship of social skills with academic achievement has been widely studied as well. There is evidence of a significant positive relationship for measures of prosocial skills, popularity, social adjustment, social competence, and cooperation scales on academic achievement from preschool through elementary school (Agostin & Bain, 1997; Doctoroff et al., 2006; Jimerson, Egeland & Teo, 1999; Malecki & Elliot, 2002; Miles & Stipek, 2006; Teo, Carlson, Mathieu, Egeland, & Sroufe, 1996; Trzesniewski et al., 2006). Evidence is mixed, however, on the effects of social skills and problem behaviors on achievement when looked at jointly. One study found that it was only measures of social skills, not problem behaviors, that maintained an effect on later academic achievement (Malecki & Elliot, 2002). Another found consistent associations between measures of social skills and problem behaviors and literacy achievement throughout elementary school (Miles & Stipeck, 2006). This last study actually found that the patterns of association differed over time, with kindergarten social skills effects diminishing over time and aggression effects increasing over time. Reasons for these contradictory results are unclear, though the authors hypothesized that social skills are particularly important in the early grades, as students are learning how to adapt to a classroom setting and are developing relationships with teachers and peers. These skills may be more important in earlier grades than in later grades.

Some research has looked into the pathways through which social skills and problem behaviors affect academic achievement. Arnold (1997) found during his observations of 74 preschool boys over a period of weeks that students with antisocial behavior problems paid less attention during class, received less help from the teacher, and in the end performed worse then their peers. Blair and Diamond (2008) maintain that self-regulation leads to positive adjustment and adaptation in the classroom (such as positive social relationships, productivity, and positive sense of self). Children with higher levels of social-emotional and regulation skills can also take advantage of learning opportunities.

Another research strand has studied measures of social-emotional skills jointly with measures of learning-related skills and regulation skills, particularly attention skills

(McGee et al., 2002; Patrick, 1997; Spira & Fischel, 2005;). Most have found it is attention and other regulatory skills, rather than social-emotional skills, that have an effect on academic achievement in elementary school (Claessens, Duncan & Engel, 2008; Duncan et al., 2007; Hinshaw, 1992a; Howse, Calkins et al., 2003; Vitaro, Brendgen, Larose & Tremblay, 2005). For example, in a study using six different datasets, Duncan and colleagues (2007) consistently found that after controlling for attention, other social-emotional skills (both social skills and problem behaviors) measured in kindergarten no longer had an effect on reading and math achievement scores in grades 3 through 5. Attention skills, on the other hand, had a consistent standardized effect of 0.1 standard deviations across all six datasets.

Other aspects of self-regulation that have been shown to significantly contribute to improved academic achievement (from emerging literacy, language and math skills in preschool to reading and math achievement in elementary school), are emotion regulation, such as effortful control (Denham et al., 2003; Eisenberg et al., 2004; Graziano et al., 2007; Howse, Lange, Farran, & Boyles, 2003; Hughes, Luo, Kwork & Lloyd, 2008; Leerkes, Paradise, O'Brien, Calkins, & Garrett, 2008; Morrison et al., 2009; Raver, Garner, & Smith-Donald, 2007), and behavior regulation, such as self-control, behavioral inhibition and activation and attention control (Blair & Razza, 2007; Fantuzzo et al., 2005; McClelland et al., 2007; McClelland, Morrison & Holmes, 2000; Ponitz et al., 2009; von Suchodoletzt, Trommsdorff, Heikamp, Wieber, & Gollwitzer, 2009). Early self-regulation skills also have shown positive long-lasting effects on later achievement. For example, one study found that higher behavior regulation skills of kindergarten students increased the achievement gap between children with low regulation skills through second grade, with the gap persisting through sixth grade (McClelland et al., 2006).

A number of reasons explain why poor self regulation skills could lead to social-emotional problems in the classroom. Poor regulatory skills limit a child's ability to navigate and effectively manage the classroom environment and take advantage of learning opportunities (Graziano et al., 2007; Howse et al., 2003; Morgan, Farkas, Tufis & Sperling, 2008; Rothbart & Bates, 1998). Children lacking regulation skills are more likely to act out, behave aggressively, resist following rules and requests of others, all of which are thought of as behavior problems (Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002; Graziano et al., 2007; Hamre & Pianta, 2001; Hughes, White, Sharpen, & Dunn, 2000; Ladd & Burgess, 1999; Morrison et al., 2009; Raver, 2004; Shields et al., 2001). Self-regulation in the classroom is demonstrated by a variety of skills and positive behaviors, such as the ability to remember and follow directions, make and follow through on plans, adapt responses when working on problems, and inhibit inappropriate responses while activating appropriate responses, such as raising one's hand to answer a question, or ignoring a misbehaving neighbor when doing individual seat work (Dowsett & Livesey, 2000; Gathercole & Pickering, 2000; Kail, 2003; Ladd & Burgess, 1999; McClelland & Morrison, 2003; Miles & Stipek, 2006, Morrison et al., 2009; Rimm-Kaufman et al., 2000; Rueda et al., 2005; Wentzel, 1993; Wilding, 2005; Zelazo, Müller, Frye, & Stuart, 2003).

Most of the research presented above looked at the effects of measures of social-emotional and regulation skills on achievement in the same year or the effects of early skills over the space of a few years. Those who have examined effects over longer periods of time tend to find a diminishing effect over time (for example, see Miles & Stipeck, 2006, or Gutman et al., 2003 for a counterexample). I found only one research study that examined the effect of repeated measures of social skills over time on achievement trajectories, finding the effect of social skills remained fairly stable over time (Gutman et al., 2003). This continual relationship could be due, in part, to the reciprocal effects of social-emotional and regulation skills and achievement. Evidence for a reciprocal model of effects of social competence and academic achievement has been found in elementary school students (Mercer & DeRosier, 2008; Miles & Stipeck, 2006; Morgan et al., 2008; Rabiner & Coie, 2000; Rowe & Rowe, 1999; Trzesniewski et al., 2006). Reciprocal relationships could lead to downward or upward spiraling patterns of development over the years (Buyse, Verschueren, Doemen, Van Damme, & Maes, 2008). For example, attention problems have been shown to lead to poorer reading achievement, which, in turn, could lead to a child's becoming more frustrated during the school day, increasing his or her problem of paying attention, which, in turn, could lead to lower reading achievement. While this might not be the case for all children, this potential pattern demonstrates the importance of intervening as soon as possible for children with regulation and social-emotional problems.

Another way of looking at the relationship between social-emotional and regulatory skills and academic achievement is to examine the intervention literature. Reviews of the social-emotional learning (SEL) intervention literature found that these interventions are effective at improving the social and emotional functioning of the skills they are teaching (Payton et al., 2008). Evidence of the effectiveness of these SEL programs for improving elementary student's academic achievement, however, is mixed. Researchers at CASEL performed a review of all types of SEL program and found a meta-analysis effect size of .28 at post-test on academic achievement outcomes which was maintained at follow-up. Effects were larger for interventions targeting children with problems – .43 and .67 at post-test and follow up, respectively. Other meta-analytic reviews, however, have found much smaller effects on achievement (e.g. Wilson, Gottfredson, Najaka, 2001).

Recent randomized trials of SEL interventions have found that while the interventions have strongly affected the SEL skills being targeted, few, if any, have affected academic achievement (Barnett et al., 2008; Domitrovich et al., 2007; Jones et al., in review; Raver et al., 2008). Evidence from Head Start REDI, which combined an SEL intervention (PATHS) with an academic intervention (High/Scope or Creative Curriculum) in Head Start preschool classrooms, did find significant but small differences between children in the intervention and control groups on vocabulary and emergent literacy measures (Bierman et al., 2008). Limited effects on achievement have also been found from the 4Rs intervention, an elementary school SEL intervention, though only for children with the greatest risk of behavior problems who have received two years of 4Rs (Jones et al., under review). Such limited evidence of improved academic achievement from interventions targeting social-emotional skill is discouraging. This lack of impact could be because the interventions are not targeting the skills that are most closely linked to academic achievement (such as attention and other regulatory skills), or it could be the interventions are not intensive enough to affect the desired change in achievement. More evidence, however, is needed to add to the mixture of results described here in order to understand how and why SEL interventions affect achievement.

Summary

This chapter began outlines the constructs within the domains of health and social-emotional skills. Health has typically been divided into two types: (1) physical health, encompassing areas such as chronic illnesses, obesity, malnutrition, and general health status; and (2) mental health, referring to the child's overall well-being in developmental, social, emotional and cognitive areas. Mental health constructs relate closely to constructs within the domain of social-emotional skills. These can be divided into three general areas: social skills, problem behaviors, and learning-related skills. Regulation skills are closely related to these three areas, and many researchers believe they form the foundation of social-emotional skills and behaviors. A variety of measures from both the health and social-emotional skills domains have been shown to be related to academic achievement. Historically, researchers have tended to focus on the effect of a single construct on achievement, but recently more researchers have begun to look at multiple related constructs within each domain in an effort to determine more precisely which are related to achievement. Initial evidence suggests that attention and other learning-related skills have the strongest and longest lasting effects on achievement.

While there appears to be a substantial evidence base looking at the relationship between health and social-emotion and regulation skills on academic achievement, more needs to be done. First, only a handful of studies have investigated social skills, behavior problems, and regulation skills jointly, and while these have consistently found that regulation skills are most closely related to achievement, additional studies are needed to verify this result in order to gain wider support for this conclusion from the education research community. Additionally, more needs to be done to look into the continuing effect over time of time-varying measures of health and social-emotional skills on achievement rather than just the long-term effects of early measures. The next part of this dissertation is a secondary data analysis that addresses both of these shortcomings in the literature.

Chapter 4
Methodology

Data Source

This dissertation uses data from the Early Childhood Longitudinal Study - Kindergarten class of 1998-1999 (ECLS-K), sponsored by the National Center for Education Statistics of the U. S. Department of Education (USDE NCES, 2006). The ECLS-K implemented a multistage probability sample design to select a nationally representative sample of kindergarten children in 1998. The ECLS-K sampled roughly 22,000 kindergarten children from over 1000 sampled schools, and by 2004 had tracked the early school experiences of these children from kindergarten through fifth grade. The ECLS-K will continue to track this cohort of children through high school. ECLS–K data allow researchers to chart the academic achievement of school-age children with some precision and to examine how academic achievement varies for students from different family and social backgrounds attending schools with different demographic and organizational characteristics. The large, nationally representative samples of schools and children in ECLS-K ensure sufficient variability in backgrounds, abilities, and social competence of children—as well as in school and teacher practices—to allow comparison of academic achievement over time of different children from different school and classroom settings.

Data collection for ECLS-K occurred for all the children in the fall and spring of kindergarten, spring of first grade, spring of third grade, and spring of fifth grade, with an additional collection period in the fall of first grade for a sub-sample of around 30 percent of the original cohort. I used data from all but the fall of first grade round for my analyses. For each round, the ECLS-K gathered questionnaire data from parents, teachers, and school administrators, and gave direct cognitive assessments of the children in reading, math, and general knowledge/science skills.

Study Sample

Because I was interested in tracking the learning trajectories of children from kindergarten through fifth grade, I limited my sample to children who were in the initial ECLS-K sample and were still in the study by the spring 2004, when the fifth-grade data collection round occurred. Children who missed a round of data collection between kindergarten and fifth grade were retained in the sample, as their learning trajectories could still be calculated in the statistical model I used in my data analyses. I also only included children who attended mainstream education settings and were in fifth-grade classrooms by the Spring 2004 data collection period, therefore excluding children in special education classrooms or those who were held back a grade. My final sample included 11,613 children. The attrition leading to the loss of almost half of the original sample occurred primarily due to children moving from their original schools, which I describe in more detail in the following section.

Missing Data

Two main causes of missing data in the ECLS-K were mover attrition and non-response. Of the nationally representative sample of kindergarteners, the ECLS-K followed all of the children who remained in the same school, but only followed a sub-sample of children who transferred schools in first grade, third grade, and fifth grade. At each new data collection point after kindergarten, each child was labeled as (1) a stayer in the same school, (2) a mover, flagged for follow up, (3) a mover, not flagged. Most of the movers who were flagged for follow up were found, and those who were not found were considered non-responders. Some of the stayers and flagged movers at each time point were non-respondents or individuals who returned only partially completed surveys. Table 4.1 summarizes the different causes of missing data.

These various reasons for missing data could lead to a number of potential problems for my analyses. First, if movers were different from stayers, mover attrition could lead to significant bias in estimates and could even lead to a portion of the sample being lost. The ECLS-K attempted to reduce the bias from mover attrition through a complex sampling strategy of movers and the use of weights, as described in the following section on mover attrition. Secondly, non-response could also lead to biased

Table 4.1
Description of eligible and ineligible children in ECLS-K

Eligible Children		Ineligible Children
Stayer	**Mover, Flagged to Follow**	**Mover, Not Flagged**
• responder	• responder	--
• non-responder	• non-responder (not found or failed to respond)	--

estimates if the responders were significantly different from non-responders. I address this issue in the section on non-response. I conclude by describing the strategy used to impute missing non-response data.

Mover Attrition

The original ECLS-K sample was nationally representative of the population of kindergarten students in American in 1998-1999. At each new data collection period, the ECLS-K followed all of the children who remained in the same school, but only followed only a sub-sample of children who transferred schools in first grade, third grade, or fifth grade. The ECLS-K chose a sampling strategy at each grade level designed to ensure no loss of representivity in the sample over time. In the spring of first grade, a random sample of 50% of kindergarten schools were flagged to have all of their movers followed, with priority given to students who had been included in the fall of first-grade data collection point. In the spring of third grade, movers in a new random sample of 50% of first grade schools were flagged to have their students followed. Additionally, to ensure that the number of children for certain subgroups of interest, such as language-minority children, did not drop too low, all mover children in certain subgroups were automatically flagged to be followed. In an effort to reduce costs, smaller sub-sampling rates were used in fifth grade, with an attempt to maximize sampling of children with full longitudinal data, and to over-sample movers from subgroups of interest.

Table 4.2 shows the percent of movers, percent of movers flagged to be followed, and the total number of eligible students for each grade. By fifth grade, only slightly

Table 4.2 Mover attrition over time in ECLS-K					
	Fall K	**Spring K**	**Spring Grade 1**	**Spring Grade 3**	**Spring Grade 5**
% Movers		6%	25.7%	42.3%	40.1%
% Movers Flagged		100%	47.8%	53.9%	41.9%
# Eligible	21,356	21,941	17,652	16,829	12,126

more than half of the original sample were still eligible and being followed. Because the ECLS-K mainly used a random sampling of schools to flag movers for follow-up, I was not concerned that any subpopulation was lost from the data.

Mover attrition, however, could still lead to biased estimates if children who moved schools were different from children who stayed in the same school, or if the group of mover children flagged to be followed were different from the group of mover children not flagged for follow up. The ECLS-K was able to calculate weights, based on each child's probability of selection at each round, to compensate for imbalances due to mover attrition. These weights allowed for accurate estimations of population means, but were only available for children for whom ECLS-K had complete data at each of the five major data collection periods. Because I included children in my sample who were non-responders at the first and third grade time points, I did not use the ECLS-K provided weights in my analyses. By not weighting, my overall estimates of population means were not generalizable to the 1998-1999 kindergarten cohort of children in America[2].

Although I did not use weights in my analyses, I was not overly concerned with bias from mover attrition for a number of reasons. First, because the sampling of movers was performed, for the most part, randomly at the school level, I assumed the ineligible (not flagged) children were missing at random. As I accounted for school level clustering, as well as the children's new schools, in my analytic models, the estimates and standard errors should not be biased. As a precaution, I also compared the different groups of children across a set of covariates that were correlated with achievement and that ECLS-

[2] For methods of estimating population means using covariates instead of weighting, see Firth & Bennett, 1998.

K documentation had shown might be different by groups (Tourangeau, Lê, Nord, & Hausken, 2005). I first calculated the mean differences[3] between all children who were ineligible and those who were eligible in fifth grade, which are presented in Table 4.3. I found that ineligible children tended to have slightly lower teacher ratings on the social rating scale, were more likely to be Black, but less likely to be Hispanic (due to the inclusion of all language-minority students in follow-up samples), and were more likely to come from lower SES homes and more single-parent homes. All of these differences,

Table 4.3
Unstandardized mean differences and standard errors for eligible and ineligible ECLS-K children in fifth grade

	mean	*(se)*	
Approaches to Learning (0-4)	0.1023	(0.0109)	***
Self-Control (0-4)	0.0847	(0.0104)	***
Interpersonal Skills (0-4)	0.0751	(0.0105)	***
Externalizing Problems (0-4)	-0.1113	(0.0109)	***
Internalizing Problems (0-4)	-0.0404	(0.0083)	***
Health Status (1-5)	0.0218	(0.0129)	
Child has Disability	-0.0132	(0.0056)	*
BMI	0.0284	(0.0166)	
Birth Weight, oz	1.3230	(0.4985)	**
Male	-0.0152	(0.0078)	*
Black	-0.0210	(0.0048)	***
Hispanic	0.0437	(0.0054)	***
Asian	0.0093	(0.0040)	*
Other	-0.0020	(0.0368)	
Socio-Economic Status	0.0313	(0.0115)	**
Non-English Spoken in Home	0.0394	(0.0059)	***
Child Repeated Kindergarten	-0.0085	(0.0034)	*
Single Parent	-0.0530	(0.0067)	***
Parent Education Expectations (1-5)	0.0689	(0.0213)	**
Unsafe Neighborhood	-0.0149	(0.0068)	*
Full Day Kindergarten	-0.0012	(0.0039)	
Kindergarten Class Size	0.0765	(0.0496)	
Kindergarten Class Behavior (1-5)	0.0349	(0.0118)	**

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

[3] Mean differences were calculated using HLM, in order to account for the clustering of children in schools.

though statistically significant, were fairly small. I also compared movers flagged for follow up to movers not flagged for follow up at each time point. Results from these mean comparisons are presented in Table A.1 in the Appendix. As expected, I found few differences at first grade (when the sample was fully random), with more differences appearing as the sampling strategy began to include all children from specific subgroups of children. None of these differences, however, were substantially large.

In order to ensure my model estimates were not biased due to mover attrition, I controlled for all measures that had differences. I also tested whether movers had a significant impact on my growth models by including an indicator of whether or not children had been movers in my models of reading and math achievement. Once I accounted for all control variables in the models, I found that the indicator measure for being a mover was not significant on the intercept or growth terms.

Non-Response

Another potential source of bias in my model estimates was non-response. Although every effort was made to collect complete data at each round, a significant amount of data was missing in the ECLS-K, both from non-completion of surveys and item non-response, where items were left blank on returned questionnaires. Table 4.4 lists the overall completion rates of child assessments and parent, teacher, and school administrators' questionnaires for all eligible children at each round of data collection. Overall, completion rates were quite good for a large-scale longitudinal survey. The

Table 4.4
Completion rates over time in ECLS-K

	Fall K	Spring K	Spring Grade 1	Spring Grade 3	Spring Grade 5
Child Assessment	89.8	88.3	91.8	86.1	93.6
Parent Interview	84.7	83.8	85.8	80.3	90.7
Teacher Questionnaire	91.3	86.0	83.5	70.0	90.6
School Administrator Questionnaire		85.4	81.4	73.3	89.6

highest completion rates were for child assessments, averaging 90% across all data collection periods[4]. Overall, third grade had the lowest level of completion rates, and fifth grade had the highest.

Non-response also arose from items left blank in returned questionnaires. Overall, I found that most of the variables I used in my analyses had less than 15% missing data (from item non-response or non-completion of survey). Exceptions included the teacher social skills rating scale in third grade, and most teacher measures in grades 1 and 3, which had between 15-30% missing data. I did not include any variables in my analyses with more than 30% missing data. While the amount of non-response appears large, these amounts are comparable to, or better than many large-scale longitudinal surveys.

Due to my concern that responders and non-responders were not similar, I compared the mean differences between these two groups on the same covariates as those used above. I found there were only some small, but significant differences between responders and non-responders on the covariates tested, as seen in Table A.3 of the Appendix. Non-responders at all but the fifth grade time point were not different from responders on initial reading and math achievement. There were also no differences on most measures of the child's background and family, nor on classroom measures. Males were more likely to be non-responders, as were children with disabilities. Non-responders were also more likely to receive lower teacher ratings on the social rating scale. In an attempt to reduce bias from non-response, I included the imbalanced covariates in my growth models. I also included these covariates when imputing missing data, in order to improve the accuracy of the imputation of my measures of interest.

Missing Data Imputation

In order to retain each child in the sample for my analysis, I imputed values for missing child, parent, teacher and school data using multiple imputation. Rather than replacing missing values with a single value, multiple imputation replaced missing values with a vector of plausible values. The variability in this set of plausible values represented the uncertainty about the actual value. I used the SAS multiple imputation

[4] Completion rates for child assessments of stayers and flagged movers who were found were very similar – in the mid to upper 90's with the overall average completion rate being brought down by flagged movers who could not be found.

procedure to impute the ECLS-K data. For the imputation process, I used the Markov chain Monte Carlo (MCMC) method, which assumed an arbitrary pattern of missingness and multivariate normality of the missing data (Schafer, 1997).

Due to the multi-level nature of the ECLS-K data, I imputed the data separately by level. First I created 5 imputed datasets of child and parent items. This included the direct assessments of children's achievement, teacher and parent ratings of social-emotional skills, and measures of the child, family, home, and neighborhood as reported by a parent in the parent survey. I then created five imputed datasets of teacher items based on the teacher survey. Finally, I created five imputed datasets of school-level items from the school administrator questionnaire. For both the teacher and administrator imputations, I included aggregate measures of student achievement to improve the quality of the imputation.

For my final analyses, I ran each of my models five times, once for each imputed dataset, and then combined the results. I calculated the overall estimates by taking the mean of the results, Q, from the five imputed datasets. To compute the variance and standard error of each result (such as a coefficient in my growth model), I followed Rubin's (1987) method of combining the within-imputation and between-imputation variance, as follows:

Within-Imputation Variance, U

$$\bar{U} = \frac{1}{m}\sum_{j=1}^{m} \hat{U}_j$$

where U is the standard error of the estimate and m is the number of imputations.

Between-Imputation Variance, B

$$B = \frac{1}{m-1}\sum_{j=1}^{m} \left(\hat{Q}_j - \bar{Q}\right)^2$$

Where Q is the estimate of interest and m is the number of imputations

Overall Total Variance, T

$$T = U + \left(1 + \frac{1}{m}\right)B$$

Where U is the within-imputation variance and B is the between-imputation variance.

Measures

In this section I outline the measures used for my analyses. I begin by describing the cognitive assessment measures used, followed by measures of children's social-emotional skills and health. I then outline the set of child, family, classroom, teacher, and school measures included as covariates in my statistical models.

Measures of Child Academic Achievement

The ECLS–K used an adaptive testing procedure to administer reading, math, and either general knowledge or science achievement tests to students at each round of data collection. These assessments were designed not only to measure a student's knowledge at each data collection period, but also to measure academic growth over the course of the study. To ensure accurate measurement of ability over time and to reduce floor and ceiling effects, ECLS-K designed its tests to be adaptive and used Bayesian approaches to Item Response Theory (IRT) to create a common scale across rounds of test administration (Pollack, Atkins-Burnett, Rock, & Weiss, 2005). Each test administration round used three test forms of differing difficulty levels, and students were given a short routing test to determine the appropriate test form each should receive. IRT methods used data from all the tests across all rounds to create a common scale of ability estimates. As a result, gains in this scale at different points in time were comparable. The IRT ability scale was built using common test items found across forms and across rounds. The IRT ability estimates had high internal item-consistency reliabilities, hovering around 0.95 at each round.

The ECLS-K derived three types of scores from the IRT ability estimates used in my growth models: scale scores, standardized scores, and proficiency scores. Scale scores were nonlinear transformations of ability status at a given point in time. Because the scores were derived using IRT methods, point gains in ability at different points in time were comparable. However, without knowing what specific items were used, observed gaps in scale scores do not easily denote what, or how much, a child knew. An additional concern with scale scores was that they were based on an arbitrary scale created with respect to the specific set of items used in the ECLS-K. Additions of different items with varying difficulties could change the scale and thus the size of the gaps (Reardon, 2008).

A final concern was the differences in the variability of scale scores over time. In the ECLS-K, students' standard deviations of reading and math achievement doubled from the beginning of kindergarten to the end of third grade, and then decreased in fifth grade. This increasing then decreasing variance made it difficult to compare gaps at different grades, as differences in point scores at third grade were not equivalent to point differences in kindergarten.

The second type of score, standardized scores, were computed in such a way as to remove the problem of increasing variability over time. Standardized scores provided estimates of achievement relative to the population as a whole. These scores were rescaled to a mean of 50 and standard deviation of 10 at each round, allowing comparisons of individual or group performance relative to others at a given cross-section of time. While not ideal for studying gaps over time, increases in standardized scores for an individual reflected an increase in the relative ranking of that individual's test scores with respect to other individuals over time.

The ECLS-K also reported each child's highest level of proficiency achieved in reading and math at each grade level. The kindergarten and first grade reading assessments focused on basic skills (print familiarity, letter recognition, beginning and ending sounds, creating rhyming words, "sight" word recognition), vocabulary, and comprehension (listening comprehension and understanding words in context). The third and fifth grade assessments measured phonemic awareness, single word decoding, vocabulary and passage comprehension. These sets of skills were divided into 9 proficiency levels, as follows: (1) *letter recognition*: identifying upper- and lower-case letters of the alphabet by name; (2) *beginning sounds*: associating letters with sounds at the beginning of words; (3) *ending sounds*: associating letters with sounds at the end of words; (4) *sight words*: recognizing words by sight; (5) *comprehension of words in context*: understanding words in context; (6) *literal inference*: making inferences using cues directly stated with key words in the text; (7) *extrapolation*: identifying clues used to make inferences; (8) *evaluation*: demonstrating understanding of author's craft and making connections between a problem in the narrative and similar life problems; and (9) *evaluating nonfiction*: comprehension of biographical and expository text (Pollack et al., 2005).

The kindergarten and first-grade mathematics assessment focused on conceptual knowledge, procedural knowledge, and problem solving. The third and fifth grade assessments measured number sense, properties, and operations; measurement; geometry and spatial sense; data analysis, statistics, and probability; and pattern, algebra, and functions. These sets of skills were divided into 9 proficiency levels, as follows: (1) *number and shape*: identifying some one-digit numerals, recognizing geometric shapes, and one-to-one counting up to 10 objects; (2) *relative size*: reading all one-digit numerals, counting beyond 10, recognizing a sequence of patterns, and using nonstandard units of length to compare the size of objects; (3) *ordinality and sequence*: reading two-digit numerals, recognizing the next number in a sequence, identifying the ordinal position of an object, and solving a simple word problem; (4) *addition and subtraction*: solving simple addition and subtraction problems; (5) *multiplication and division*: solving simple multiplication and division problems and recognizing more complex number patterns; (6) *place value*: demonstrating understanding of place value in integers to the hundreds place; (7) *rate and measurement*: using knowledge of measurement and rate to solve word problems; (8) *fractions*: solving problems using fractions; and (9) *area and volume*: solving word problems involving area and volume (Pollack et al., 2005).

I focused the bulk of my analyses and discussion on the IRT scale scores, as I was interested in children's growth in reading and math achievement over time. However, due to concerns that effect sizes over time from scale scores could be biased due to the increasing variability in the scores over time, I also ran my growth model using standardized scores as the outcome to test the validity of the scale score results. Lastly, I ran an ordinal HLM model using proficiency scores as the outcome to better understand what gaps might mean in terms of differences in children's learning.

Measures of Child Social-emotional Skills

As I outlined in the review of the literature, measures of social-emotional skills cover a wide range of areas—social skills, problem behaviors, learning-related skills, self-regulation—which many researchers believe underlie some of the more overt social-emotional skills. The ECLS-K contained teacher, parent and child ratings of these first three areas, based on the Social Rating Scale (SRS), a variation of the Social Skills

Rating Scale (SSRS) (Gresham & Elliot, 1990). The scales of the SRS included approaches to learning, self-control, interpersonal skills, externalizing problems, and internalizing problems.

In this section I first describe who rated the children on these scales, the correlations of the scales across raters, how researchers dealt with discrepant information from multiple raters in the past, and how I dealt with the multiple raters in the ECLS-K data. I then describe in detail the teacher SRS scales and the regulatory behaviors factor I created from four of the five SRS scales. Lastly, I discuss my examination of the stability of the SRS scales and regulatory behaviors over time, to determine whether to include these measures as time-varying covariates or stable factors in my growth model.

<u>Description of Teacher, Parent, and Child Ratings</u>

The ECLS-K obtained ratings on social behaviors from teachers, parents and children. Teachers rated children's social behaviors at each round of data collection, while parents only rated children in the kindergarten and first grade rounds, and the children rated themselves in the third and fifth grade rounds. Five scales were created from the questions asked in the teacher and parent ratings: approaches to learning, self-control, interpersonal skills, internalizing problems, and externalizing problems. Children only rated themselves on interpersonal skills and problem behaviors. Correlations between informants was low, with parent-teacher correlations hovering between .1 and .2 on most of the scales at each time point. Correlations were marginally higher for externalizing problem behaviors. Teacher-child correlations were around .1 for interpersonal skills, .18 for internalizing problem behaviors, and .35 for externalizing problem behaviors. These correlations were consistent with across-rater correlations found in previous research looking at internalizing and externalizing problem behaviors (Achenbach, McConaughy & Howell, 1987; Brown, Wissow, Godamski, Zachary, & Bartlett, 2006).

The reasons for such large discrepancies across raters in this dataset and others stem from multiple sources. First, ratings are dependent on different motivations, thresholds, and perceptions of an informant concerning the problem behavior, to whom the informant might be comparing the child, as well as an informant's relationship and emotions towards the child being rated (De Los Reyes & Kazdin, 2005; Richters, 1992; Youngstrom, Loeber & Stouthamer-Loeber, 2000). Reports also could differ based on ethnic and cultural backgrounds of informants (Jensen et al., 1999). The context where the child was observed could provide another source of variance in reports from different raters. Children could demonstrate different behaviors at home than at school, so different informants could reasonably provide different reports. In fact, some researchers have found that children's behavior at home might not be as overt as their behavior at school (Achenbach, Dumenci & Rescorla, 2002; Verhulst et al., 2003).

Faced with such discrepant reports of social behavior, researchers are forced to decide either which informant to use or how to combine information from multiple informants. This choice has a serious impact on identifying children with behavior and regulation problems and thus on the true effect of these problems on learning (De Los Reyes & Kazdin, 2005). Most studies rely on a single informant, and a long-standing debate exists in the literature about which informant is more accurate and reliable. Some researchers argue that parents have more opportunities for observing their child and can provide the most accurate reports (Glascoe, 2000), while others argue that parent reports are more likely to be biased, and teacher reports are more appropriate, especially as teachers are reporting from the context where children are learning – school (Fendrich, Johnson, Wislar & Nageotte, 1999; Hinshaw, Han, Erhardt, & Huber-Dressler, 1992, US DHHS 2001).

Alternatively, there has been a move towards combining information from multiple informants in an effort to find the core underlying skills and remove sources of error (Achenbach et al., 1987, Essex et al., 2006, Kraemer et al., 2003, Ollendick & Hersen, 1993, Youngstrom et al., 2003). However, there is no set system for combining multiple informants' data, with some researchers reporting each informant's data simultaneously while others aggregate data from all informants into one measure (Kraemer et al., 2003, Offord et al., 1996). These simple combinations of rater information, however, provide no guarantee of a reduction in error and bias. One promising new method for combining data from multiple sources uses principal component analysis in an effort to capture the child's actual characteristics, while removing the bias and error due to different informants' contexts and perspectives

(Kraemer et al., 2003). This method requires data from at least three distinct informants in at least two settings, allowing error from the context and informant to be identified and removed.

While I would ideally like to combine information from multiple sources, the ECLS-K dataset has only two measures of each behavior at any given point, which is not enough to use methods such as principal component analysis to remove sources of error. Instead, I used only the teacher ratings of social behavior in my model. One reason for this choice was the teacher ratings were the only ones gathered for each round of data collection from kindergarten to fifth grade. Additionally, the teacher measure was based on observations of children in the classroom context, which was where the children were being taught reading and math, my outcomes of interest. It is important to note that the bias and measurement error inherent in teacher ratings of social behavior could affect the results. Although I cannot be certain of how the results might be biased, I would expect results based on these ratings to be an underestimate of the true effect.

Details on the Teacher Social Rating Scale

In this section I describe the rating scale used by teachers in more detail. The teacher Social Rating Scale used a frequency scale (1=never to 4=very often) to rate how often children exhibited certain social skills and problem behaviors. Twenty-four items were used in the SRS in kindergarten and first grade and two new items were added to the third and fifth grade questionnaires to increase variance and reliability. Researchers who have used the SRS have described it as a measure of teachers perceptions of children's social-emotional competence (e.g. Gershoff, 2003; NICHD ECCRN, 2004; Wilson, Pianta & Stuhlman, 2007), as a measures of learning-related skills or behavioral regulation (e.g. McClelland & Morrison, 2002), and as a measure of self-regulation or behaviors indicative of self-regulation (e.g. Barnett et al., 2008; Blair & Razza, 2007).

NCES does not provide users with the individual SRS items, providing only five factor analysis scales. These scales were calculated by taking the mean of the ratings of all the items in the scale. The scales and available descriptions of the associated items are as follows:

Approaches to Learning: This scale measured behaviors that affected the ease with which children could benefit from the learning environment. It included six items that rated the child's attentiveness, task persistence, eagerness to learn, learning independence, flexibility, and organization. This scale is often referred to in the literature as attention and learning-related behaviors. Duncan and colleagues (2007) found this scale performed very much like attention scales from other datasets.

Self-Control: This scale included four items that indicated the child's ability to control behavior by respecting the property rights of others, controlling temper, accepting peer ideas for group activities, and responding appropriately to pressure from peers. This scale has also been referred to in the literature as self-regulation.

Interpersonal Skills: This scale included five items that rated the child's skill in forming and maintaining friendships, getting along with people who are different, comforting or helping other children, expressing feelings, ideas, and opinions in positive ways, and showing sensitivity to the feelings of others. This scale has also been referred to in the literature as social skills, social competence, and social functioning.

Externalizing Problems: This scale measured acting out behavior and included five items that rated the frequency with which a child argued, fought, became angry, acted impulsively, and disturbed ongoing activities.

Internalizing Problems: This scale included four items that asked about the apparent presence of anxiety, loneliness, low self-esteem, and sadness in the child.

Split-half reliabilities for each of the scales was fairly high, hovering around .9 for the approaches to learning, interpersonal skills, and externalizing problem behavior scales at each collection period, and around .8 for the self-control and internalizing problem behavior scales at each collection period. A split-half reliability is a measure of internal consistency, meaning items reflecting the same constructs yield similar results. Although these SRS scales were internally consistent at each point in time, there was no way of knowing how reliable and consistent the ratings were over time.

I had two major concerns with the teacher SRS ratings. The first was that four of the five scales were highly correlated within a single teacher. The second concern was that different teachers rated children over time, and the over-time correlations were small. Both of these issues could lead to error or bias in the model coefficient estimates. I address each of these problems below.

Correlations Between Teacher SRS Scales

Teacher ratings of approaches to learning, self-control, interpersonal skills, and externalizing problem behaviors were highly correlated, with correlations ranging from .6 to .8 across the scales. Such similar ratings by a single rater have been described as the halo effect – where teachers assign similar ratings on different aspects of behavior (Buyse, Verschueren, Doumen, Van Damme, & Maes, 2008; Mashburn, Hamre, Downer, & Pianta, 2006). This halo effect could occur for a number of reasons. First, the underlying constructs of social skills, self-control, and problem behaviors are also highly correlated. While the literature suggests these areas are correlated, prior research using other assessments has not shown such high correlations as those seen here (e.g., McClelland, Cameron, Wanless, & Murray,2007). A second explanation for the high correlations across scales could be that teachers assign students ratings based on their perceptions of how good a student each child is. If this were the case, it would suggest that the SRS is not really measuring the social skills and problem behaviors constructs.

Outwardly observed behaviors could stem from a number of different underlying causes. For example, a student with limited English proficiency may be marked by the teacher as having high externalizing problems because the child appears to be acting out a lot. That child, however, may not actually have had poor externalizing problems, but rather may have been struggling and acting out due to problems understanding English. Teachers' perceptions of student behavior can also have an effect on how the teacher interacts with the student and the relationship developed between the student and teacher, which, in turn, could affect both achievement and later behaviors. While I do not directly test whether this transactional relationship is occurring in my achievement models, this could be a possible pathway to explain a relationship between the SRS scales and achievement.

A final possibility for such high correlations across teacher scales is that there might be a latent construct that underlies all four of these scales, and the teacher ratings are representative of that underlying construct. As discussed in the earlier review of the literature, self-regulation is closely associated with social-emotional skills, and many researchers believe regulation forms the foundation for the more overt social-emotional skills that were being rated by teachers in the SRS. In fact, a number of other researchers who have used the SRS or the SSRS (Social Skills Rating Scale, on which the SRS was based) have referred to these entire scales as measures of self-regulation or behavior regulation (e.g. Barnett et al., 2008; Blair & Razza, 2007; McClelland & Morrison, 2002).

Using the term self-regulation to describe the entire SRS scale, however, has its challenges. Historically, self-regulation was most commonly rated by scales such as the SRS. Over the years, however, the understanding of what self-regulation is, and how it should be measured has changed dramatically. While behavior ratings scales are still used today, more researchers are turning to more direct measures of self-regulation. The difficulty, however, lies in the fact that researchers use the term same term—regulation—to describe both direct measures of regulation skills and observer ratings of behavior. For example, Morrison and colleagues have developed a direct assessment of executive function skills called Head, Toes, Knees Shoulders (HTKS) which they validated by comparing to teacher ratings on the SSRS and the Child's Behavior Rating Scale—ratings also commonly referred to as behavior regulation (McClelland, Cameron, Wanless & Murray, 2007; Ponitz, et. al., 2008; Ponitz, McClelland, Matthew, & Morrison, 2009). The correlations found between direct and indirect measures suggest that observer ratings are viable substitutes for measuring a child's regulation. They are not, however, identical, and such broad usage of common terminology referring to potentially different underlying constructs makes it difficult for researchers to clearly identify and state what it is that is being measured as well as what it is that is having an affect on academic outcomes.

I began this section by describing the high correlations between four of the five teacher SRS scales. I outlined a number of possible reasons for these high correlations, and the challenges of knowing what it really is that is being measured and the difficulties of choosing a term to describe the combined SRS scale. With such high correlations between these measures, researchers have two choices for how to include these measures in models. The first is to include each of the individual scales simultaneously, and the second is the combine the highly correlated scales into a single factor. While most researchers using the SRS data have included each of the individual scales in the model at the same time, doing so can lead to problems with multicollinearity. Multicollinearity leads to unpredictable fluctuations in coefficient estimates of highly correlated measures included simultaneously in a single model, making it difficult to know if the estimates are reliable. While I would ideally like to include all of the individual scales in my model in order to determine if one scale has a stronger independent effect on achievement than the others, I was concerned that multicollinearity would make any results from those models suspect. I thus decided to combine the approaches to learning (attention), self-control, interpersonal skills, and externalizing problems scales into one factor. Factor analysis revealed one principal component explaining 71-75% of the variance in the subscales at each time point. Factor loadings all hovered around 0.8-0.9 for each of the scales, so I took the mean value of the four scales to create the factor score.

The major difficulty in using a factor of the four SRS scales is in determining what it is that factor represents. As outlined above, the factor could be a representation of how good a student the child is (based on teacher's perceptions and the halo effect), or it could be that the four scales all have the same underlying latent construct. Unfortunately, it is impossible to know what this factor is really measuring. Attaching a label to this factor is thus challenging. As mentioned earlier, similar labels are often used to depict substantially different measurements, leading to confusion concerning what constructs are really being represented. While I would argue, based on evidence of the strong correlation between regulation and social-emotional skills, that regulation may be the construct underlying the four SRS scales, and other researchers have used the term regulation to refer to the SRS, I am wary of using that term to label the factor, as the term regulation has recently come to represent a much more precise set of characteristics and skills than the behaviors being measured by the teachers in the SRS. In order to signify that these four SRS factors may be based on regulation while recognizing also that they are also teacher perceptions of the child's behaviors, I have chosen to label this factor regulatory behaviors.

Correlation of Teacher Ratings over Time

Another problem with the social-emotional measures in the ECLS-K was that different teachers rated children at each collection period, and correlations over time were fairly low.[5] As one might expect, ratings at the beginning and end of kindergarten, which were largely rated by the same teacher, had decent correlations, near 0.6 for all but externalizing problems, which had a correlation of 0.7. Across all other time points, the correlations for self-control and interpersonal skills ranged from 0.3 to 0.4, internalizing problem behaviors from 0.2 to 0.3, and externalizing problem behaviors from 0.4 to 0.5. Two main possibilities exist for such low correlations across time. The first, as discussed earlier, is that different teachers could have different understandings and perceptions of what behavior problems are, and could also have different personal biases, thus making the scales un-comparable across raters and time. The other possibility is that children's behaviors fluctuate over time and the low correlations are a reflection of this natural fluctuation. Unfortunately, it is impossible to determine if the differences in scores across time are due to changes in the children's behavior or due to differences in the raters and context.

The ambiguity in the reason for low correlations over time for repeated measures leads to the question of whether the SRS scales should be treated as time-varying covariates in the model, or whether they should be averaged and considered stable characteristics of the child. The child development literature was divided on this issue, with many claiming middle childhood as a time of great change in social-emotional skills, while others claimed that children's functioning, or at least their rank order, was fairly stable over time (see, for example, Hinshaw, 2002; Lahey, McBurnett, & Loeber, 2000; Patterson, Forgatch, Yoerger, & Stoolmiller, 1998). Measurement error makes it difficult to determine stability based solely on rater scales. As noted earlier, correlations over time were fairly low, but the reasons for this (whether it was due to different raters and systematic error, or due to actual fluctuations in the measure) were unclear.

---

[5] This problem is not unique to ECLS-K data, but occurs in most longitudinal datasets measuring social-emotional skills over time.

Measurement models, however, can provide a rough estimate of the amount of variability in the SRS measures that was due to fluctuations over-time. A proper measurement model uses the individual items that make up a scale, items that the ECLS-K does not provide in their dataset. However, because I combined four of these scales to create the regulatory behaviors factor, I was able to mimic a measurement model, using the externalizing, interpersonal, self-control and approaches to learning scales as items.

In a simple measurement model, an observed score, X, would be composed of the true score, T, and error, *e*, which is: $X=T+e$. The true score can thus be thought of as the observed score minus error, or the expected value of X. The reliability of the observed score is the ratio of true score variance to observed score variance. Put into hierarchical linear model terminology, a measurement model for regulatory behaviors would include 3 levels: Level 1 would be the items (the four SRS scales), which would be nested in regulatory behaviors over time (Level 2), which would be nested in children (Level 3). This is specified as follows:

Level 1

$$Y=\pi_0 + e$$

Level 2

$$\pi_0 = \beta_{00} + \beta_{01}(\text{linear}) + r_0$$

Level 3

$$\beta_{00} = \gamma_{000} + u_{01}$$
$$\beta_{01} = \gamma_{010} + u_{01}$$

In this model, π_0 represented the true regulatory behaviors score at time t. The term *e* in the model represented error, which could be composed of both systematic and random error. Systematic error, which I assumed makes up the largest portion of the error in the model, could be error due to rater bias and contextual effects.

To determine the stability of the regulatory behaviors scale over time, I first determined the sources of variability. Total variability in the intercept (regulatory behaviors scale) in this three-level model was composed of (1) between-child variability, $\tau_{between}$, (2) within-child / between-time variability, τ_{within}, and (3) error, σ^2. Once I removed the variability due to error, the remaining variability of the true scores was between-child and within-child variability. I could then estimate the stability of regulatory behaviors over time by looking at the percent of variability in regulatory behaviors due to differences between children: $\tau_{between}/(\tau_{between} + \tau_{within}) = 0.39611 / (0.39611 + 0.19851) = .6662$. In other words, 66.6% of the variability in regulatory behaviors true scores was between children, leaving only 33.4% over time. This indicated that substantial stability existed over time and supported a decision to use the average regulatory behaviors true score over time as a stable factor in the model.

I also looked at the stability over time of the five original SRS scales. Because I did not have the individual items forming these measures, I used a 2 level HLM model of timepoints within children. The variability of the error term σ^2 was now composed of all within-child variability – from both systematic/random error and variability due to time. It was impossible to disentangle these two sources of variability, though a quick and dirty method would be to use the reliability of the intercept to estimate the amount of error in the model. As mentioned earlier, reliability is the ratio of true-score variance to observed-score variance, so a high reliability indicates a smaller amount of random error. Using this reliability, I attempted to remove the random error component by using only the 'reliable' proportion of variance in σ^2, calculated by multiplying σ^2 by the reliability. The proportion of between-child variability, using the quick and dirty method, would then be $\tau_0/(\tau_0+\text{reliability}*\sigma^2)$.

Table 4.5 contains the stability estimates, as represented by the percent of variability between children for each of the SRS scales, using the standard method (not removing error), and the estimated, or quick-and-dirty, method. Only two of the original five scales, externalizing problems and approaches to learning had over 50% of their

Table 4.5
Stability of Social Rating Scale factors over time

	Intern. Problems	Extern. Problems	Interp. Skills	Self-Control	Approaches to Learning	Regulatory Behaviors Factor
% variability between children, standard method	28.8%	54.3%	40.8%	43.2%	53.8%	41.6%
% variability between children, estimated method	37.7%	58.2%	47.1%	49.0%	57.7%	66.6%

variability between children, even after estimating without the error. The combination of scales in the regulatory behaviors factor was by far the most stable, indicating that while behaviors themselves seemed to fluctuate over time, the underlying trait of regulation appeared to be fairly stable. Although there was a basis in the developmental literature for assuming stability in regulatory skills, as described earlier, many researchers have assumed these traits were changing and developing over the middle childhood years. Because of this, I used regulatory behaviors as both a stable and time-varying measure in my models, to determine if substantial differences existed in estimates of effects on achievement for the different specifications.

Summary of Social-Emotional Skill Measures Used in Data Analyses

In my data analyses, I used the regulatory behaviors factor and the internalizing problems scale as measures of social-emotional skills. Based on the measurement model, evidence showed that regulatory behaviors could be a stable trait. Because including this factor as a stable trait was a fairly novel approach, I chose to report my results using the factor as both a stable trait (averaged over time) and as a time-varying trait from the true-score estimates from the measurement model. I also included internalizing problems as a time-varying covariate in the model.

For both of these time-varying measures, I accumulated the scores from kindergarten through time t-1 at each time point. I accumulated the measures over time based on the assumption that the effect of an earlier measure of behavior was not completely mediated by later measures, but could still have a unique effect on achievement. I used a series of 2-level (children nested within schools) hierarchical linear models of reading achievement to test this assumption. I ran a model with all prior measures of regulatory behaviors on the end-of-first-grade reading achievement, another for end-of-third-grade achievement, and another for end-of-fifth-grade achievement. I found that even after controlling for the most recent measurement of regulatory behaviors, all but the kindergarten entry measures remained significant in each model. Figure 4.1 shows a graphical depiction of these series of models over time. I found that the effects of the earlier measures did slowly diminish over time, but remained significant even after including more recent measures in the model.

In order to account for multiple measures at a given time point in a longitudinal analysis, I could not simply include different measures of regulatory behaviors on each growth rate. Rather, I created a time-varying covariate, accumulating all prior measures for each time point, as depicted in Table 4.6[6]. For example, the effect of regulatory behaviors on third grade math achievement was the combined effect of regulatory behaviors from the fall and spring of kindergarten and the spring of first grade.

Figure 4.1
Total effect size of regulatory behaviors on reading achievement from four separate HLM models

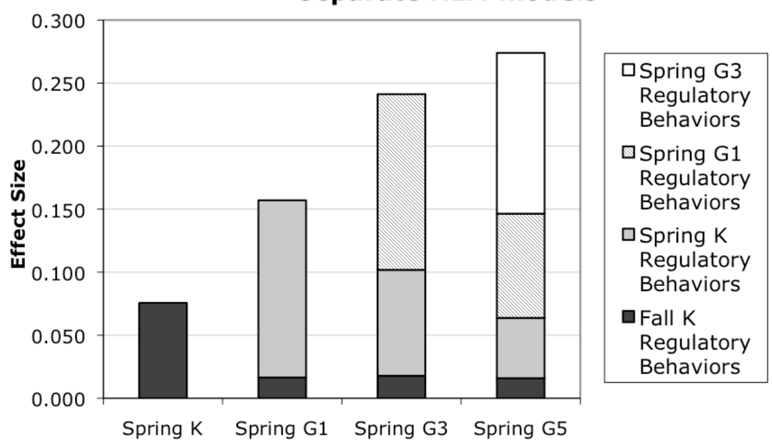


Table 4.6
Method for accumulating repeated child measures over time when creating a time-varying covariate

Fall K	Fall K
Spring K	Fall K
Spring G1	Fall K + Spring K
Spring G3	Fall K + Spring K + Spring G1
Spring G5	Fall K + Spring K + Spring G1 + Spring G3

[6] Including a single measure in my model that was comprised of the summed values of the repeated measures over time was similar to including all the individual measures in the 2-level achievement models mentioned earlier.

Measures of Child Health

In addition to these measures of social-emotional skills, I also included several measures to represent a child's health. The first was a parental rating of a child's general health, reported on a 5-point scale from "poor" to "excellent." This health status is better thought of as parents' perception of their child's health, as there was no way to know what "poor" or "good" health meant to each parent. Health status, in fact, has many of the measurement problems associated with ratings of social-emotional skills. The one difference is that, for the most part, the same parent rates the child over time. Surprisingly, however, this did not lead to high correlations over time. Correlations from year to year hovered around .4. I also looked at whether some children had more variability in scores over time than others. I found that approximately 50% of the children had moderate to large fluctuations in health status over time, 24% had small fluctuations over time, and the rest were consistently rated as having 'very good' to 'excellent' health. Because the majority of children's health fluctuated (whether due to actual fluctuations or measurement error, I cannot be certain), I included health as a cumulative time-varying covariate in my models.

Parents also reported, at each collection round, whether or not their child had a disability. This could be a physical disability such as vision, hearing, speech, or mobility; a mental health disability, such as activity or learning problems; or the use of child-received special-education services[7]. I included this measure as a cumulative time-varying covariate. I included two measures representing the child's health at birth: whether they were premature and their birth weight in ounces. I also included the child's body mass index (BMI), which was the ratio of weight to height, and was reported each round. BMI was highly correlated over time and was based on a measurement model like those run for the SRS scales. Eighty-one percent of the variability was between children (before adjusting out any error). I thus chose to include BMI as a stable trait, and used the average BMI in my model. Lastly, I included three health-related measures. The first was

[7] It is important to note that there could be considerable inconsistency concerning who was and wasn't marked as disabled. First, parents could mark their child as having a mental health disability without having received a clinical diagnosis for the problem. Secondly, states and school districts have different standards and requirements for who can receive special education services (an indicator of having a disability), suggesting that being labeled by as disabled might be inconsistent by state. Model coefficients for this measure should thus be interpreted with caution.

whether or not the child had health insurance. The second was whether or not the child had visited either a doctor or dentist within the past year. I included both of these measures as cumulative time-varying covariates in my model. The final measure was a parent rating of food insecurity, as reported in kindergarten.

Measures of the Child, Family, and Home

Along with the health and social-emotional skills measures, I included a large set of controls in my models in order to improve my causal claims. For covariates describing the child and family, I used measures taken from the parent questionnaire describing child and family background characteristics, family structure, parent investments in the child, parental beliefs, and home environment. Child background characteristics included were gender, age, and race (White, Black, Hispanic, Asian, and other minority). I also included whether the child came from a single-parent home, the family's socio-economic status (an ECLS-K computed measure of parent income, education and occupational prestige), parent mental health problems, and parent expectations of how much schooling they thought their child should receive (six categories from less than high school to PH.D or other higher degree). To describe the home environment, I included how often the parents did educational and cognitively stimulating activities in the home, whether arguing and other challenging interactions took place in the home, how often the parents read to the child in the home, whether the parents had rules for watching TV, and whether the child's bedtime varied. Lastly, I included the safety of the child's neighborhood, which was the average score of parent reports of amount of garbage, drug use, burglaries, violent crimes, and vacant houses in the area around the home.

Sameroff (2000) proposes it is the number of risk factors, rather than the nature of risk factors, that best determines outcomes (see also Sameroff & Fiese, 2000). A single risk factor is not enough to change developmental trajectories, but having multiple risk factors can lead to large deflections, especially since risk factors tend to cluster in individuals (Gutman et al., 2003). Sameroff and colleagues proposed the use of a risk index – a sum of the risk factors affecting the child – in developmental models. Numerous studies have tested the soundness of this strategy and have found that while including each individual measure in the model is more predictive, risk indexes do account for significant variability in the outcomes. I used a risk index in my model in order to test more easily the interaction of health and social-emotional skills with one measure of the home environment on reading and math achievement. I included the following risk factors in my index: 1) SES, 2) single-parent household, 3) parent mental health, 4) parental expectations for educational attainment, 5) home environment, 6) educational activities in the home, 7) amount of reading in the home, and 8) rules for bedtime. Using a measurement model, I found that 72% of the variability in this index was between children, computed without adjusting out error, indicating that the child's family and home environment was fairly stable over time. I thus included the average home risk index in the model.

Measures of the Teacher, Classroom, and School

At the end of each school year, teachers reported on their characteristics, beliefs, and practices as well as the composition of their classroom. While these types of measures are rather coarse, they can be useful as a set of controls in statistical analyses of students' academic achievement outcomes. To represent the context where children learn, I included measures of the classroom composition, class behavior, parent participation, and teacher characteristics. Classroom composition measures included the percent of minority students in the class, the percent of children with disabilities in the class, and the percent of students who were reading below grade level in the class. I included a teacher rating of overall class behavior, which was rated on a five-point scale from "misbehaves very frequently" to "behaves exceptionally well." Parent participation measures included were teacher reports of the percent of parents who volunteered in the classroom, whether the teacher contacted the child's parents to discuss the child (for both good and bad reasons), and whether the child's own parent had ever volunteered in the classroom. Teacher characteristics included were teacher race, age, number of years teaching, certification, and whether they had a masters degree or higher.

School data were obtained from questionnaires filled out by administrators at the end of each school year. Administrators reported on school structural features, school composition, and school policy. I only included measures in my analyses that had been measured during each round of data collection. The type and location of schools was captured by whether the school was public or private, the school's urbanicity (urban, suburban, rural), and the region of the country where the school was located (Northeast, South, West, Midwest). I also included measures of total school enrollment, percentage of students eligible for free or reduced lunch, percentage of students reading below grade level, number of nurses in the school, number of special education staff in the school, and whether or not having a quiet and orderly environment was a priority for the school.

Data Analysis Procedures

The overarching question of this dissertation was how students' health and social-emotional skills affect their reading and mathematics achievement during their elementary school years. To explore this association, I used a three-level hierarchical cross-classified longitudinal model (HCM) with either math or reading achievement as the dependent variable, health and social-emotional skills measures as predictors, and a wide range of child, family, home, and school measures as controls (Meyers & Beretvas, 2006; Raudenbush & Bryk, 2002, pg. 373). I describe this model in more detail in the following section.

HCM of Math and Reading Achievement

I modeled math and reading achievement from kindergarten to fifth grade with a hierarchical cross-classified model (HCM). The HCM model allowed me to take advantage of the nested sampling structure of ECLS-K as I modeled reading and math achievement over time for children cross-classified by schools. By fifth grade, over 50% of the children in ECLS-K had moved to a different school. A regular 3-level longitudinal model limited analysis to those children who remained in the same school across all grades. However, children who moved might not share the same characteristics as those who remained in the same school, potentially biasing any results from such an analysis. With HCM I accounted for the changing schools some children move through during elementary school, retaining these children in my dataset. The HCM model allowed children to be cross-classified with two or more schools during their elementary school years.

The first step in building this model was determining the appropriate growth curve that would best fit the IRT scale score data. Upon inspecting overall growth curves and sampling individual growth curves, I identified four different growth rates: one from the fall to spring of kindergarten (K), one from the spring of kindergarten to the spring of first grade (G1), one from the spring of first grade to the spring of third grade (G3), and one from the spring of third grade to the spring of fifth grade (G5). I consequently decided to use a piecewise model to capture these different growth rates (Raudenbush & Bryk, 2002, pg. 178). A piecewise model is a growth model that is made up of a series of linear pieces, in this case the periods of time between each data collection of the ECLS-K.

I ran the HCM model four times for four separate outcomes. The first two outcomes were the IRT scale scores of math and reading. I also ran two models using the IRT standardized scores for math and reading to check the validity of the scale score model.

Level 1 Model: Time Points

To calculate the achievement status for child j at time t, I first specified the level 1 model, corresponding to time points within children. I coded the four growth terms in the model (K, G1, G3, and G5) so that the intercept in the model (π_{0jt}) corresponded to achievement in the fall of kindergarten for child j at time t. Typically, a level 1 model was modeled solely with the growth rates, but I also included a number of time-varying covariates at level 1, labeled in the model as TVC. Time-varying covariates were level 1 predictors that varied over time, thus explaining variations in the outcome (Hong & Raudenbush, 2008; Raudenbush & Bryk, 2002, pg. 179). In my model these were regulatory behaviors, internalizing problems, health status, disability status, whether or not the child visited the doctor or dentist over the past year, and whether or not the child had health insurance, as well as all teacher/classroom measures. I also included the interaction of three of these time-varying covariates (regulatory behaviors, health status, and disability status) with the individual growth rates (TVC*K, TVC*G1, TVC*G3, and TVC*G5). Including these interaction terms allowed these three time-varying covariates to have different effect sizes for each grade, rather than a single rate across all time. The level 1 model was specified as follows:

$$
\begin{aligned}
\text{Achievement Status}_{tjk} = \pi_{0jk} \\
&+ \pi_{1jk}(K) + \pi_{2jk}(G1) + \pi_{3jk}(G3) + \pi_{4jk}(G5) \\
&+ \pi_{ujk}(TVC) && \text{for u=5 to 23} \\
&+ \pi_{vjk}(TVC*K) && \text{for v=24 to 26} \\
&+ \pi_{wjk}(TVC*G1) && \text{for w=27 to 29} \\
&+ \pi_{xjk}(TVC*G3) && \text{for x=30 to 32} \\
&+ \pi_{yjk}(TVC*G5) && \text{for y=33 to 35} \\
&+ e_{tjk}
\end{aligned}
$$

$$\text{where } e_{ijt} \sim N(0, \sigma^2)$$

In this model, Achievement Status_{ijt} is the achievement in reading or math at time t for student j in school k. The coefficient π_{0jk} is the mean achievement status at the beginning of kindergarten for child j in school k. The next four π coefficients refer to the growth in achievement for kindergarten (K), kindergarten to first grade (G1), first to third grade (G3), and third to fifth grade (G5) for child j in school k. Table 4.7 shows the vector of values each growth rate takes to calculate achievement status at time t. For example, achievement status at the end of first grade is the sum of initial status, kindergarten growth, and first grade growth. The remaining π coefficients indicate the increase in achievement for child j and school k for that time-varying covariate. For example, for interpersonal skills, the achievement status for child j in school k would increase by π_{6jk} at each time point with a 1 unit difference in interpersonal skills, controlling for all other terms in the model. For regulatory behaviors, which had both the time-varying covariate and an interaction term between the TVC and the piecewise growth rates, the effect would be the combination of those terms. For example, for a 1 unit difference in regulatory behaviors at the beginning of kindergarten, the achievement status at the end

Table 4.7
Growth rate values used to calculate achievement status at each time point

	Kindergarten Growth Rate	Grade 1 Growth Rate	Grade 3 Growth Rate	Grade 5 Growth Rate
Fall K	0	0	0	0
Spring K	1	0	0	0
Spring G1	1	1	0	0
Spring G3	1	1	1	0
Spring G5	1	1	1	1

of kindergarten increases by both π_{5jk} (the coefficient for the TVC) and π_{24jk} (the coefficient for TVC * K growth rate) for child j in school k, controlling for all other terms in the model. The term e_{tjk} represents the deviation at time t for student j in school k from the predicted achievement status. The error term is normally distributed with a mean of 0 and a variance of σ^2.

Level 2 Model: Row (Child) and Column (School)

I next modeled the intercept and four growth rates for reading and math achievement with measures of child, family, teachers and schools, using deviance test statistics to create the best possible model. I allowed the intercept, first grade, third grade, and fifth grade growth rates to vary by child, and the kindergarten growth rate to vary by school. In the model this is specified with the terms b and c. Here, b_{00j} is the random effect associated with child j on initial status. The term b_{i0j} for i=1,2,3,4 is the random effect associated with child j on the math or reading learning rate for the given grade. c_{10k} is the random school effect, or the expected deflection to the growth curve associated with encountering school k. $D_{hjk} = 1$ if student j encountered school k at time h, 0 otherwise. I allowed the error term for kindergarten growth to vary only across schools, because the model would not support allowing school-level error to vary across more than one growth rate.

The level 2 models were specified as follows:

$$\pi_{0jk} = \theta_0 + b_{00j}$$
$$+ \beta_{0x}(\text{Child \& Family measures}) \qquad \text{for x=1 to 16}$$
$$+ \gamma_{0y}(\text{School measures}) \qquad \text{for y=1 to 10}$$
$$\pi_{1jk} = \theta_1 + D_{hjk}c_{10k} + \beta_{1x}(\text{Child \& Family measures}) + \gamma_{1y}(\text{School Measures})$$
$$\pi_{2jk} = \theta_2 + b_{20j} + \beta_{2x}(\text{Child \& Family measures}) + \gamma_{2y}(\text{School Measures})$$
$$\pi_{3jk} = \theta_3 + b_{30j} + \beta_{3x}(\text{Child \& Family measures}) + \gamma_{3y}(\text{School Measures})$$
$$\pi_{4jk} = \theta_4 + b_{40j} + \beta_{4x}(\text{Child \& Family measures}) + \gamma_{4y}(\text{School Measures})$$
$$\pi_{qjk} = \theta_q \qquad \text{for q=5, 6, … 35}$$

In this model, θ_0 refers to the expected math or reading achievement at the beginning of kindergarten, when all the predictors in the model are set to 0. I chose to center all continuous measures, so 0 is the mean. The terms θ_1, θ_2, θ_3, and θ_4 refer to the expected growth in reading achievement, when all predictors are 0 over the kindergarten year, first grade, first to third grade, and third to fifth grade, respectively. The remaining θ's refer to the expected increase in achievement for the given time-varying covariate when all predictors are set to 0, as specified in Table 4.8. The β's in the model are the

Table 4.8
Specification of model coefficients

Level 1 Coefficients for Child and Teacher Time Varying Covariates		**Level 2 Coefficients for Child-level and School-level Measures**	
θ_5	Regulatory Behaviors	β_{i1}	Average Regulatory Behaviors
θ_6	Internalizing Problems	β_{i2}	Birth Weight
θ_7	Health Scale	β_{i3}	Premature
θ_8	Disability Status	β_{i4}	Average BMI
θ_9	No Doctor/Dentist Visit in Past Rear	β_{i5}	Kindergarten Food Insecurity
θ_{10}	No Health Insurance	β_{i6}	Male
θ_{11}	Class Size	β_{i7}	Black
θ_{12}	% Minority in Class	β_{i8}	Hispanic
θ_{13}	% Disability in class	β_{i9}	Asian
θ_{14}	% Read below Grade Level in Class	β_{i10}	Other
θ_{15}	Teacher Rating of Class Behavior	β_{i11}	Age
θ_{16}	Time Spent in Reading Instruction	β_{i12}	No English in Home
θ_{17}	Teacher Calls Home for Good Behavior	β_{i13}	Repeat Kindergarten
θ_{18}	Teacher Calls Home for Bad Behavior	β_{i14}	Parent Chooses School
θ_{19}	Parent Volunteered in Class	β_{i15}	Behavior Skills Important for K
θ_{20}	# Volunteer Hours	β_{i16}	Average Home Risk Index
θ_{21}	Teacher of Minority Race	β_{i17}	Full Day Kindergarten
θ_{22}	Teacher Has Masters' Degree or Higher	γ_{i1}	Northeast
θ_{23}	# Years Teaching Experience	γ_{i2}	South
θ_{24}	Regulatory Behaviors * Spring K	γ_{i3}	West
θ_{25}	Health Scale * Spring K	γ_{i4}	Private School
θ_{26}	Disability Status * Spring K	γ_{i5}	School Enrollment
θ_{27}	Regulatory Behaviors * Spring G1	γ_{i6}	% students below grade level
θ_{28}	Health Scale * Spring G1	γ_{i7}	Special Ed FTE
θ_{29}	Disability Status * Spring G1	γ_{i8}	Nurse FTE
θ_{30}	Regulatory Behaviors * Spring G3	γ_{i9}	Policy for Quiet/Orderly Environment
θ_{31}	Health Scale * Spring G3	γ_{i10}	School Neighborhood Safety
θ_{32}	Disability Status * Spring G3		
θ_{33}	Regulatory Behaviors * Spring G5		
θ_{34}	Health Scale * Spring G5		
θ_{35}	Disability Status * Spring G5		

unstandardized coefficients representing the increase (or decrease) in the intercept or growth rate due to a 1 unit increase in the given child and family measures while the γ's are the coefficients for school measures, as specified in Table 4.8. The full model listing of all variables and coefficients is given in the Appendix.

Effect Size Computation

To demonstrate the size of the direct effects on reading and math achievement status and growth, I calculated the effect size, which provides researchers with a common metric for comparing effects within and across studies. Social scientists typically view a standard deviation effect greater than 0.5 as large and below 0.2 as small (Cohen, 1988). To calculate effect sizes, I used the raw standard deviation of reading or math status for children at each time point (Tate, 2000). Using the standard deviation of reading or math status for children at each time point accounted for the increase in variation in the outcome over time due to the IRT scale score. As described earlier, the standard deviations of reading and math achievement doubled from the beginning of kindergarten to the end of third grade and then decreased in fifth grade. This increase in variance makes it difficult to compare gaps in children's scores at different points in time. What appears to be an increasing gap might actually be the same size, after taking into account differences in variance over time. Using the variance of children's ability status at each time point, rather than using only the variance from kindergarten entry, standardizes the differences over time, making gaps comparable over time. Using the raw standard deviation rather than the model-based standard deviation provides a more conservative estimate

I calculated the effect size at each grade by summing the coefficients for the kindergarten entry intercept and related growth rates prior to and including time t and dividing this by the overall student standard deviation at time t. For example, the effect size for regulatory behaviors at first grade is the sum of the coefficients from the intercept (beginning of kindergarten), the kindergarten growth rate, and the first-grade growth rate, divided by the overall child standard deviation at first grade. I also computed the achievement gap in months of learning, calculated by dividing the coefficients prior to and including time t and dividing this by the overall reading growth rate. Once again, using the overall growth rate is a conservative estimate of size, and can be considered a lower bound to the true effect size.

Testing the Sensitivity of the HCM Reading and Math Models to Specification Error

The major concern of the reading and math hierarchical cross-classified models was that the results could have been biased due to specification error. In order to test the sensitivity of my results, I ran two additional analyses. The first analysis replaced the reading and math scale scores with standardized scores to determine if the growing gaps found in the IRT scale score were real or a result of the increasing variability in the scale over time. The second analysis re-ran the HCM model for math and reading scale scores, including propensity strata. Including propensity strata in the model allowed me to test for the possible impact of confounding variables. If the standardized coefficients were similar across models, I could say with some confidence that my results were robust.

<u>Standardized Score Analysis</u>

The IRT scale scores for reading and math provided by the ECLS-K were scaled specifically for use in longitudinal models. The variability in the scores, however, increased through third grade, then decreased slightly in fifth grade, making interpretations of the size of gaps over time unclear. As described, I attempted to account for this changing variability when calculating effect sizes by using the standard deviation of each time point, rather than using only the standard deviation at kindergarten entry. I also tested for the accurateness of the gaps over time by using the standardized scores provided by the ECLS-K as dependent variables in the HCM model. The standardized score, as the name suggests, is standardized at each time point to a mean of 50 and a standard deviation of 10 and represents an individual's rank order at each time point. The standardized scores are designed to compare groups against each other, rather than comparing a single group over time. If the standardized scores for a group increase (or decrease) over time, the change reflects an increase (or decrease) in the relative ranking of this group's test score with respect to that of other groups over time. If the gap for regulatory behaviors, for example, actually does increase over time, I would expect their relative rank to change as well, as reflected in the standardized score.

Propensity Score Stratification Analysis

A second concern with the reading and math HCM models was the possibility of confounding variables, which I addressed using propensity score stratification. Propensity scores are most easily described when used for experimental data with treatment and control groups. In this situation, the propensity scores refer to the child's propensity for receiving treatment. These propensity scores are calculated by predicting treatment status from a wide range of available measured covariates. Using the propensity score, the study sample can be divided into balanced strata. Each of these strata contain children from both the treatment and control groups with similar measured characteristics, who I assume have similar unmeasured characteristics.

While children do not receive any treatment in this study, I could still use propensity score stratification by determining children's propensity for having regulatory behaviors. Using a 2-level hierarchical linear model of children nested within schools, I predicted children's average regulatory behaviors using a wide range of child, family and home background, teacher, and school measures (Raudenbush & Bryk, 2002). I then split the children in the sample into five strata, based on the children's propensity scores. Using the randomization inference omnibus test[8], I found these five strata were balanced (χ^2=8.59, df=39) on a large set of child and classroom covariates (Hansen & Bowers, 2008). Complete results of the omnibus test are presented in Table A.3 in the Appendix. Having balanced strata, I subsequently reran the reading and math HCM models with four dummy variables representing the five propensity strata.

[8] The randomization inference omnibus test (Hansen & Bowers, 2008) allowed me to test for overall balance between samples on a large number of covariates, and held a number of advantages to using the more standard logistic regression technique. Logistic regression required a considerable number of cases when running analyses with a large number of covariates; otherwise the results tended to have high type I error rates (rejecting the null hypothesis when it is actually true), affecting the reported p-values. In fact, statisticians have found that considerably more than 10 times the number of cases are required for the number of variables used. Randomization inference, on the other hand, assesses the balance on individual covariates with the adjusted means between groups, and uses a weighted sum of squares of differences of means to compute the omnibus measure of balance. This method does not assume groups are sampled from a different population, and does not inflate error rates or require large numbers of cases per variable included.

Modeling Proficiency Scores

After running the HCM models for reading and math achievement, I was interested in determining what the effect sizes for my measures of interest in health and social-emotional skills meant in terms of differences in what the children knew, or could do in reading and math. The ECLS-K calculated 9 proficiency levels each to cover the reading and math scales from kindergarten through fifth grade and provided researchers with the highest proficiency in math and reading the children achieved at each grade. The proficiency levels are listed in Table 4.9 – note that they are listed from highest to lowest, or hardest to easiest. Because proficiency scores are ordinal, I used a 2-level ordinal HLM model of children nested within schools in order to estimate the effect of health and social-emotional skills on children's reading and math proficiency scores at fifth grade (Raudenbush & Bryk, 2002, pg. 317).

The ordinal model in HLM is a cumulative probability model. Predictors in the model, such as regulatory behaviors, predict the probability of students' achieving a given proficiency level. Because the model is cumulative, rather than predicting the probability that the proficiency level (R) is extrapolation (R=3), the model predicts the probability that R<=3, or in other words, that the child is proficient at extrapolation (R=3), evaluation (R=2), and evaluating nonfiction (R=1). The probability for each response, given the covariates, β, in the model, is specified in the level 1 model below. The probabilities are computed using a cumulative logic function, e.g. P'(1)/(1 - P'(1)), represented in the model by η_{mij}, where m is the ordered categories (proficiency levels) of the response variable for student i in school j.

<u>Level-1 Model</u>

$$\begin{aligned} \eta_{mij} = \beta_{0j} & \\ &+ \beta_{1j}(\text{Beg. K Achievement}) \\ &+ \beta_{xj}(\text{Child, Family \& Classroom Measures}) \qquad \text{for } x=2 \text{ to } 18 \\ &+ D_{19ij}\delta_{19j} \end{aligned}$$

<u>Level-2 Model</u>

$$\beta_{0j} = \gamma_{00} + \gamma_{0y}*(\text{School}) + u_0 \qquad \text{for } y=1 \text{ to } 11$$

$$\beta_{xj} = \gamma_{x0} \qquad \text{for } x=1 \text{ to } 18$$

$$\delta_{19j} = \delta_{19}$$

Table 4.9
Reading and math proficiency levels

Reading Proficiency Levels		Math Proficiency Levels	
R=1	Evaluating Nonfiction	**R=1**	Area and Volume
R=2	Evaluation	**R=2**	Fractions
R=3	Extrapolation	**R=3**	Rate and Measurement
R=4	Literal Inference	**R=4**	Place Value
R=5	Comprehension of Words in Context	**R=5**	Multiplication and Division
R=6	Sight Words	**R=6**	Addition and Subtraction
R=7	Ending Sounds	**R=7**	Ordinality and Sequence
R=8	Beginning Sounds	**R=8**	Relative Size
R=9	Letter Recognition	**R=9**	Number and Shape

In the level 2 model, β_{0j} is the overall level of reading or math proficiency, and it is allowed to vary randomly over schools. δ_{19j} is the threshold, which is an intercept for category m, where D_{19ij} is an indicator for category m. (i.e. D_{19ij} =1 if m=2, D_{19ij} =0 if m=1). The coefficients for each predictor, γ_{x0}, are difficult to interpret on their own. To determine the predicted probability of student achievement at the highest proficiency level for a 1 unit difference in a predictor, say regulatory behaviors, I converted the coefficient using the following equation: $1/1+\exp(\gamma_2)$, where γ_2 is the coefficient for regulatory behaviors. For each additional proficiency level, I added the threshold for that proficiency, δ, to the equation, resulting in $1/1+\exp(\gamma_2+\delta)$.

Chapter 5

Results

Overview

In this dissertation I used data from the ECLS-K to explore the effect of health and social-emotional skills measures on reading and math achievement over the elementary school years. Specifically, I addressed the following two research questions:

1) How do health and social-emotional skills affect reading and math achievement over time, controlling for other child, family, home, and school factors?
2) What specific factors are associated with low social-emotional skill levels, and how large are the effects of these factors when combined with other factors on children's academic achievement over time?

My first research question asked what the effects of health and social-emotional skills were on reading and math achievement over time, controlling for all other covariates in the model. In answering this question, I first hoped to determine what measures within these domains had the strongest effect on reading and math achievement, holding constant all other health and social-emotional skills measures, as well as all other controls in the model. Secondly, by including health and social-emotional skills measures as time-varying covariates in my model, I was able to track how the size of the effect of these measures on achievement changed over time. Historically, researchers have looked only at the long-term effects of early health or social-emotional skills on later achievement. I expected that accounting for the variability of these factors over time would lead to larger estimates of effect sizes over time on achievement than had previously been found.

To answer this first research question, I used results from hierarchical cross-classified models (HCM) with either the reading IRT scale scores or math IRT scale scores as the outcome. In these models I included the following health and social-emotional skills measures as time-varying covariates: internalizing problems, health scale, disability status, whether or not the child has health insurance, and whether or not the child visited the doctor or dentist during the past year. During data preparation I found that the majority of the variability in my regulatory behaviors factor was due to differences between children, rather than variability over time. Because including this factor as a stable trait is an uncommon approach, I present results below from two separate models, one that included regulatory behaviors as a stable trait (averaged over time) and one that included regulatory behaviors as a time-varying trait.

I also used results from an ordinal hierarchical linear model (HLM) with math and reading proficiency scores as the outcome to answer this first research question. Proficiency scores helped me better understand the differences in what children with different levels of health and social-emotional skills had learned. From this ordinal model, I calculated the probability that children with different levels of social-emotional skills will reach different levels of proficiency in math and reading.

Finally, I compared the effects size of predictors in the HCM reading and math models to those from two other analyses to test the sensitivity of my initial model, the specification of the outcome, and potential confounding variables. The first analysis replaced the reading and math scale scores with standardized scores to determine whether the growing gaps found in the IRT scale score were real or just a result of their increasing variability over time. The second analysis re-ran the HCM model for math and reading scale scores with propensity strata. Including propensity strata in the model allowed me to test for the possible impact of confounding variables. If the standardized coefficients from the original model were similar to those from these two additional analyses, I would be able to say with some confidence that my results are fairly robust to specification error.

My second research question was comprised of two parts. First, I was interested in determining what other risk factors might be affecting children with low levels of social-emotional skills, based on the regulatory behaviors factor. Evidence from the research suggested that children in poverty or from minority backgrounds were more likely to have behavior problems than their peers. I was interested in seeing if there were other risk factors from the home and school environments that were also associated with poor regulatory behaviors. To determine this, I calculated the means and standard deviations on a wide range of risk factors for three groups of students: (1) a high-risk group comprised of students ranked one standard deviation below the mean or lower on the regulatory behaviors scale; (2) a low-risk group, comprised of students ranked one standard deviation above the mean or higher on the regulatory behaviors scale; and (3) an average-risk group, comprised of all remaining students.

The second part of my final research question asked what the size of the academic achievement gap would be over time across all risk factors. To calculate this gap, I took the group means of each risk factor, based on the three regulatory risk groups described above, and plugged them into the HCM reading and math models. Doing this gave me the average academic achievement for children in the high-risk, low-risk, and average-risk groups at each time point.

In the first section of this chapter, I present the results answering my first research question. I begin by presenting the results of the reading achievement models, followed by those of the math achievement models. I then compare these results to those from the two analyses designed to test the sensitivity of the original results to specification error. The second section of this chapter presents the results for the second research question. I first present the means and standard deviations by risk group, followed by the achievement gaps between risk groups over time.

Research Question 1: Effects of Health and Social-Emotional Functioning on Reading and Math Achievement over Time

In this section I begin by presenting the effects of health and social-emotional skills on reading achievement over time. I then present the effects of these measures on math achievement over time. I end this section by presenting the results from my analyses, checking the sensitivity of the reading and math HCM models to specification error.

Reading Achievement Model Results

Results from reading achievement models are presented in Table 5.1. This table includes only measures of health and social-emotional skills and a few key measures of the child's background. Table A.4 in the Appendix contains the full results with all the variables included in the analysis. I included results from two separate models in Table 5.1, one where regulatory behaviors was included as a stable factor and another where regulatory behaviors was included as a time-varying covariate. I report the unstandardized coefficients, standard errors, and p-values from both of theses models. At the bottom of Table 5.1 I report how well this model did in predicting variability in reading achievement, by listing the proportion of variance in the random effects for the intercept and growth rates that have been explained by adding all of my predictors to the model.

Comparing the unstandardized coefficients across these two reading achievement models in Table 5.1, I found the results were fairly consistent. Because of this, I focused my discussion on results from the stable regulatory behaviors model. I found that many of the health measures had a statistically significant effect on reading achievement, and both of the social-emotional skills measures I included had a statistically significant effect on reading achievement. Of the time-varying measures of health, the scale of children's overall health (θ_7), disability status (θ_8) , and whether or not the child had health insurance (θ_{10}) were all statistically significant, with overall health having a positive effect on reading, and the other two having a negative effect. Whether or not a child visited a doctor or dentist in the past year was not statistically significant. All of the interaction terms between disability status and the growth rates (θ_{26}, θ_{29}, θ_{32}, and θ_{35}) were statistically significant, indicating that the size of the effect of disability status on achievement differed by grade. For overall health, only the interaction with kindergarten growth and third to fifth grade growth were statistically significant. Of the remaining measures of children's health included as time-invariant predictors in the reading model, the results were mixed. Only food insecurity (β_{05}) had a statistically significant negative effect on reading achievement at kindergarten entry; none had a significant effect on the growth rate over kindergarten or from kindergarten to first grade; BMI (β_{34}) had a significant positive effect and food insecurity (β_{35}) had a significant negative effect on the reading growth rate from first to third grade; and food insecurity (β_{45}) had a positive effect on the reading growth rate from third to fifth grade. The reason for the switched direction of effects for food insecurity in fifth grade is unclear.

Table 5.1
Unstandardized coefficients and standard errors for covariates of interest from HCM models predicting kindergarten to fifth grade reading achievement

	Reading IRT Scale Score *with stable regulatory behaviors*		**Reading IRT Scale Score** *with time-varying regulatory behaviors*	
	β	*SE*	*β*	*SE*
π_0: Achievement at Fall Kindergarten				
θ_0: Initial Fall K Achievement	28.422	0.573 **	28.522	0.573**
β_{01}: Average Regulatory Behaviors	1.830	0.114 **		
β_{02}: Birth Weight	0.006	0.003 ^	0.007	0.003*
β_{03}: Premature	-0.049	0.321	-0.055	0.324
β_{04}: Average BMI	-0.279	0.102 **	-0.310	0.102**
β_{05}: Food Insecurity	-0.935	0.402 *	-1.001	0.404*
β_{06}: Male	-0.298	0.218	-0.642	0.212**
β_{07}: Black	0.192	0.385	-0.107	0.386
β_{08}: Hispanic	-1.552	0.358 **	-1.540	0.360**
β_{09}: Asian	3.733	0.605 **	4.064	0.599**
β_{010}: Other	-0.150	0.475	-0.170	0.478
β_{016}: Average Home Risk Index	-0.904	0.116 **	-0.965	0.116**
π_1: Linear Growth Rate from Fall Kindergarten to Spring Kindergarten				
θ_1: Fall K to Spring K Increase in Achievement	10.464	0.431 **	10.587	0.432**
β_{11}: Average Regulatory Behaviors	1.180	0.088 **		
β_{12}: Birth Weight	0.003	0.003	0.003	0.003
β_{13}: Premature	0.008	0.258	0.001	0.258
β_{14}: Average BMI	-0.053	0.081	-0.074	0.081
β_{15}: Food Insecurity	-0.084	0.354	-0.120	0.354
β_{16}: Male	-0.037	0.170	-0.276	0.166^
β_{17}: Black	-0.775	0.317 *	-1.015	0.316**
β_{18}: Hispanic	-0.472	0.282 ^	-0.463	0.283
β_{19}: Asian	0.898	0.427 *	1.094	0.426*
β_{110}: Other	0.344	0.402	0.304	0.404
β_{116}: Average Home Risk Index	-0.092	0.089	-0.135	0.089
π_2: Linear Growth Rate from Spring Kindergarten to Spring Grade 1				
θ_2: Spring K to Spring G1 Increase in Achievement	31.532	0.516 **	31.806	0.518**
β_{21}: Average Regulatory Behaviors	2.759	0.162 **		
β_{22}: Birth Weight	0.002	0.005	0.003	0.005
β_{23}: Premature	0.027	0.408	0.036	0.411
β_{24}: Average BMI	0.050	0.147	0.014	0.148
β_{25}: Food Insecurity	-0.909	0.606	-1.025	0.602^
β_{26}: Male	0.142	0.309	-0.412	0.307

	Reading IRT Scale Score *with stable regulatory behaviors*		Reading IRT Scale Score *with time-varying regulatory behaviors*	
	β	*SE*	*β*	*SE*
β_{27}: Black	-1.639	0.552 **	-2.107	0.548**
β_{28}: Hispanic	-1.210	0.551 *	-1.232	0.552*
β_{29}: Asian	0.775	0.796	1.059	0.799
β_{210}: Other	-0.693	0.722	-0.749	0.727
β_{216}: Average Home Risk Index	-0.598	0.165 **	-0.701	0.165**
π_3: Linear Growth Rate from Spring Grade 1 to Spring Grade 3				
θ_3: Spring G1 to Spring G3 Increase in Achievement	47.564	0.669 **	47.600	0.665**
β_{31}: Average Regulatory Behaviors	1.107	0.197 **		
β_{32}: Birth Weight	-0.008	0.006	-0.009	0.006
β_{33}: Premature	0.715	0.556	0.640	0.557
β_{34}: Average BMI	0.321	0.180 ^	0.343	0.181^
β_{35}: Food Insecurity	-0.869	0.712	-0.782	0.718
β_{36}: Male	0.425	0.365	0.412	0.363
β_{37}: Black	-4.935	0.689 **	-4.939	0.692**
β_{38}: Hispanic	-2.128	0.675 **	-2.196	0.679**
β_{39}: Asian	-5.182	0.891 **	-5.235	0.890**
β_{310}: Other	-3.801	0.985 **	-3.846	0.987**
β_{316}: Average Home Risk Index	-1.228	0.196 **	-1.243	0.197**
π_4: Linear Growth Rate from Spring Grade 3 to Spring Grade 5				
θ_4: Spring G3 to Spring G5 Increase in Achievement	20.127	0.520 **	20.046	0.523**
β_{41}: Average Regulatory Behaviors	-0.880	0.159 **		
β_{42}: Birth Weight	0.001	0.005	0.000	0.005
β_{43}: Premature	-0.266	0.417	-0.246	0.417
β_{44}: Average BMI	0.223	0.150	0.229	0.150
β_{45}: Food Insecurity	0.894	0.748	0.940	0.752
β_{46}: Male	0.569	0.312 ^	0.851	0.311**
β_{47}: Black	1.107	0.610 ^	1.489	0.614*
β_{48}: Hispanic	1.474	0.494 **	1.509	0.496**
β_{49}: Asian	0.692	0.712	0.407	0.711
β_{410}: Other	1.569	0.795 *	1.559	0.802^
β_{416}: Average Home Risk Index	0.201	0.162	0.260	0.162
π_5 - π_{35}: Time-Varying Covariates				
θ_5: Regulatory Behaviors			1.505	0.104**
θ_6: Internalizing Problems	-0.385	0.057 **	-0.337	0.058**
θ_7: Overall Health Scale	0.326	0.101 **	0.354	0.101**
θ_8: Disability Status	-0.714	0.303 *	-0.693	0.304*

	Reading IRT Scale Score *with stable regulatory behaviors*		Reading IRT Scale Score *with time-varying regulatory behaviors*	
	β	*SE*	*β*	*SE*
θ_9: No Doctor/Dentist Visit in past year	0.335	0.158 *	0.333	0.159*
θ_{10}: No Health Insurance	-0.065	0.205	-0.066	0.206
θ_{24}: Regulatory Behaviors* K Growth Rate			0.961	0.086**
θ_{27}: Regulatory Behaviors* K-G1 Growth Rate			-0.134	0.096
θ_{30}: Regulatory Behaviors* Spring G3 Growth Rate			-0.377	0.084**
θ_{33}: Regulatory Behaviors* G3-G5 Growth Rate			-0.639	0.053**
θ_{25}: Health Scale * K Growth Rate	0.114	0.088	0.137	0.088
θ_{28}: Health Scale * K-G1 Growth Rate	0.230	0.158	0.238	0.158
θ_{31}: Health Scale * G1-G3 Growth Rate	-0.066	0.165	-0.093	0.165
θ_{34}: Health Scale * G3-G5 Growth Rate	-0.278	0.086 **	-0.296	0.087**
θ_{26}: Disability Status * K Growth Rate	-0.736	0.237 **	-0.747	0.239**
θ_{29}: Disability Status * K-G1 Growth Rate	-1.089	0.451 *	-1.141	0.449*
θ_{32}: Disability Status * G1-G3 Growth Rate	1.105	0.513 *	1.061	0.515*
θ_{35}: Disability Status * G3-G5 Growth Rate	0.362	0.256	0.378	0.258
% Variance in Error and Random Effects Explained by Model				
b_{00}: Fall Kindergarten Intercept (Level 2) Random Child Effect	23.35%		22.69%	
b_{20}: Spring K to Spring Growth Rate Random Child Effect	14.01%		12.98%	
b_{30}: Spring G1 to Spring G3 Growth Rate Random Child Effect	4497%		3.96%	
b_{40}: Spring G3 to Spring G5 Growth Rate Random Child Effect	0.00%		0.00%	
c_{10}: Fall K to Spring K Growth Rate Random School Effect	30.24%		28.86%	
e: Level 1 Error	3.39%		2.94%	

~ $p < .10$ * $p < .05$ ** $p < .01$
Note: Coefficients in this table are unstandardized.

Both internalizing problems and regulatory behaviors had a statistically significant effect on reading achievement. The internalizing problems had a negative effect on reading achievement, while the effect of regulatory behaviors (average or time-varying) on achievement was positive. The effect of average regulatory behaviors on reading achievement was statistically significant at kindergarten entry (b_{01}=1.830) as well as on each of the growth rates, from fall to spring of kindergarten (b_{11}=1.180), from kindergarten to first grade (b_{21}=2.759), from first to third grade (b_{31}=1.107), and from third to fifth grade (b_{41}=-0.880). While I might be tempted to say that the largest effect was on the kindergarten to first grade growth rate, and a substantial negative effect was evident on the fifth grade growth rate, it is important to remember that the variability in the reading achievement scale score increases through third grade and then decreases in fifth grade. The size of these unstandardized coefficients should consequently not be compared directly over time without taking into account the increasing variability in the outcome.

In order to compare the size of model coefficients across predictors as well as over time, I calculated standardized effect sizes for each grade measured by the ECLS-K. To do this, I first calculated the total difference in achievement points at each grade by adding the unstandardized coefficients on the growth rates to the unstandardized coefficient on the intercept. For example, the total point difference for a one standard deviation difference in regulatory behaviors at first grade is the point difference at kindergarten entry plus the point difference on the end of kindergarten and end of first grade growth rates. I then converted these total point differences into standardized effect sizes by dividing the total point difference at each time point by the raw standard deviation of reading achievement at that time point.

In Figure 5.1, I graphed the standardized effect sizes on reading achievement at each time point for measures of health and social-emotional skills that had a statistically significant effect on reading achievement over time. Each line on the graph represents the effect size for a single measure, holding all other covariates in the model constant. Table 5.2 presents the effect size at each point in time in two ways: (1) the effect in points on the reading IRT scale score, and (2) the standardized effect size. In this table, I include only those health and social-emotional skills measures that had a statistically significant effect on reading achievement over time. I also include, for comparison purposes, the effect size of three measures that previous research has shown to have substantial effects on reading achievement: gender, black race/ethnicity, and average home risk index. It is important to note that the risk index includes measures of socio-economic status, family

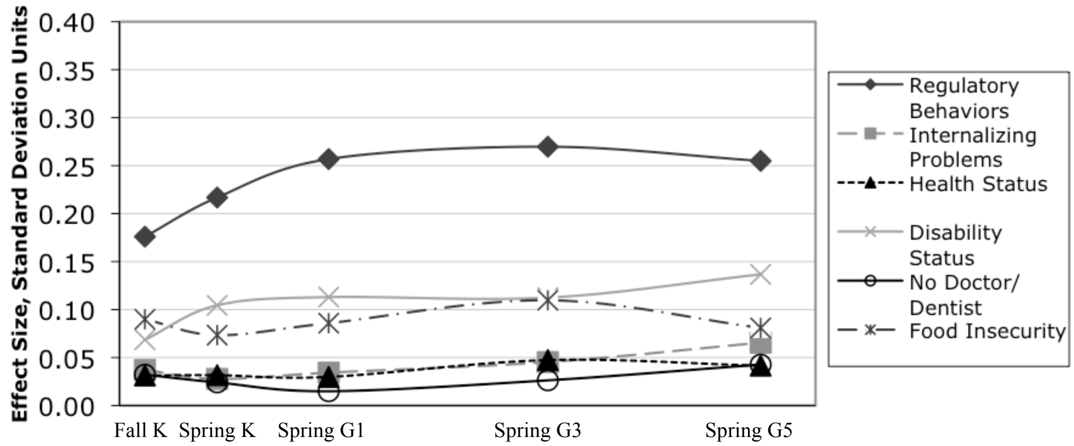
Table 5.2
Effect sizes over time of health and social-emotional skills measures and other key child characteristics on reading achievement

		Fall K	Spring K	Spring G1	Spring G3	Spring G5
Average Regulatory Behaviors	*Points**	1.830	3.010	5.769	6.876	5.996
	*SD^*	0.176	0.217	0.257	0.270	0.255
Regulatory Behaviors, Time-Varying	*Points*	1.505	2.466	4.665	5.868	5.268
	SD	0.145	0.178	0.208	0.230	0.224
Internalizing Problems	*Points*	-0.385	-0.385	-0.770	-1.155	-1.540
	SD	-0.037	-0.028	-0.034	-0.045	-0.065
Overall Health Scale	*Points*	0.326	0.439	0.670	1.207	0.977
	SD	0.031	0.032	0.030	0.047	0.042
Disability Status	*Points*	-0.714	-1.450	-2.539	-2.867	-3.215
	SD	-0.069	-0.104	-0.113	-0.112	-0.137
No Doctor/Dentist Visit	*Points*	0.335	0.335	0.335	0.669	1.004
	SD	0.032	0.024	0.015	0.026	0.043
Food Insecurity	*Points*	-0.935	-1.019	-1.928	-2.797	-1.902
	SD	-0.090	-0.073	-0.086	-0.110	-0.081
Home Risk Index	*Points*	-0.904	-0.996	-1.594	-2.822	-2.622
	SD	-0.087	-0.072	-0.071	-0.111	-0.111
Male	*Points*	-0.298	-0.335	-0.193	0.231	0.800
	SD	-0.029	-0.024	-0.009	0.009	0.034
Black	*Points*	0.192	-0.583	-2.222	-7.157	-6.049
	SD	0.019	-0.042	-0.099	-0.281	-0.257

* Points at each time point have been accumulated by adding coefficients from the intercept and all growth rates prior to and including a given time point.

^ Standard deviation effect sizes calculated by dividing total points at each time point by the raw standard deviation of reading achievement for that time point.

Figure 5.1
The standardized effect size of health and social-emotional skills measures on reading achievement from kindergarten through fifth grade



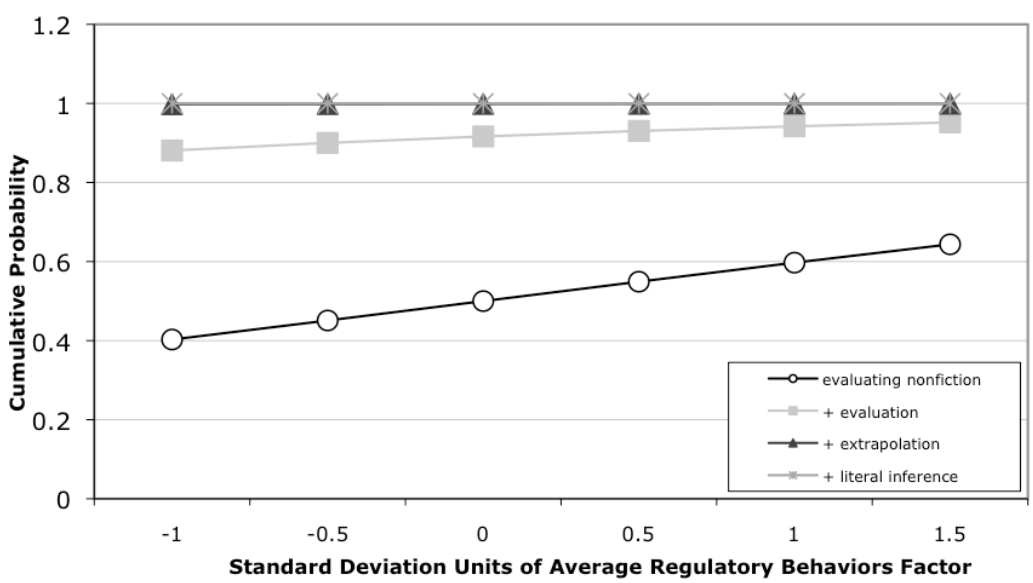
structure, home environment, parent characteristics, and parent practices and beliefs.

Social-Emotional Skills. As seen in Table 5.2, a child's regulatory behaviors had a much larger effect size on reading achievement at kindergarten entry than all the other health and social-emotional skills measures, with the exception of food insecurity. The regulatory behaviors effect size was also much larger than that of being black, male, or having a bad home-risk index. At kindergarten entry, average regulatory behaviors had an effect of 0.176 standard deviations (sd) on reading achievement, holding all other covariates constant at 0. This means that for a 1 standard deviation difference in regulatory behaviors, a 0.176 sd gap exists in reading achievement. Tracking the growth in the effect size of regulatory behaviors on reading achievement over time, as seen in Figure 5.1, I found the effect size increased over kindergarten and through first grade before leveling off through fifth grade, ending with a substantial effect on reading of 0.255 sd. The effect size of the time-varying regulatory behaviors factor on reading achievement followed the same pattern as the stable factor, but was slightly smaller, with children entering school with a 0.145 standard deviation reading achievement gap, which grew to 0.224 sd by fifth grade, holding all other covariates constant. The only other factor with an effect of this size on reading achievement by fifth grade was being black (e.s.=-0.257).

In order to understand what this gap meant in terms of what was being learned, I ran a 2-level HLM ordinal model with fifth-grade reading proficiencies as the outcome. The cumulative probability of being proficient on the highest proficiency levels is presented graphically in Figure 5.2. Children who were one standard deviation below the mean on regulatory behaviors had a 40% probability of being proficient on the highest proficiency level (evaluating non-fiction) compared to 60% of children who were one standard deviation above the mean, holding constant all other covariates in the model. The probability of reaching proficiency on the two highest proficiency levels (evaluation and evaluating non-fiction) was 88% for children who were one standard deviation below the mean and 94% for those one standard deviation above the mean of regulatory behaviors, holding constant all other covariates. The cumulative probability for all the other proficiencies was similar across levels of regulatory behaviors.

Turning to the other measure of social-emotional skills, I found, as seen in Figure 5.1, that the effect size of internalizing problems on reading achievement was much smaller than that of regulatory behaviors, but similar to those of health at each time point. The effect size of internalizing problems on reading achievement in the fall of

Figure 5.2
Predicted cumulative probability of fifth grade reading proficiency by average regulatory behaviors



kindergarten was -0.037 sd. The size of internalizing problems remained fairly steady, with minor fluctuations over time, ending fifth grade with an effect size of -0.065.

Health. The health factor with the largest and most consistent effect on reading achievement over time was food insecurity. Children in homes with high food insecurity began school -0.090 sd below their peers, and this gap was basically maintained, with some fluctuations, through fifth grade. Disability status (whether or not children had a physical or mental health disability) also had a substantial effect on reading achievement. Although the reading gap at kindergarten entry for children with a disability was only -0.069 sd, this gap increased each subsequent year, ending in fifth grade with a gap of -0.137 sd. The effects of overall health and health insurance on reading were both quite small and were basically maintained, with some small fluctuations, over time. The point in time with the strongest effects on reading for the health factors occurred at kindergarten entry, and these effects were then basically maintained in each subsequent grade.

Other Risk Factors. One of the interesting results from this model was that the unstandardized coefficients for being male and of being black were not statistically significant at kindergarten entry, as seen in Table 5.1. The coefficient for being black was statistically significant on each of the subsequent growth rates, though the size of the effect on reading achievement was not substantial until third grade (-0.281 sd), as seen in Table 5.2. The coefficient for being male, on the other hand, was only significant on the third to fifth grade growth rate. I went back and re-ran the reading model, leaving out all health and social-emotional skills measures, to see if the unstandardized coefficients for being male, being black, or of other control measures had been mediated by my variables of interest. Full mediation occurs if the addition of health and social-emotional skills measures led to the coefficients for being male or black (or any of the other covariates) no longer being significantly different than zero. Partial mediation occurs when the size of the coefficients drop, but remain significant.

I found that the size of the unstandardized coefficient for gender was fully mediated in kindergarten and first grade by the addition of health and social-emotional skills measures to the model. In other words, after accounting for children’s health and social-emotional skills, gender was no longer important in predicting reading achievement in kindergarten or first grade. The effect of being black was fully mediated at the beginning of kindergarten and was reduced in subsequent grades. The effect of being Hispanic was slightly mediated in all grades, as was the effect of the home/family risk index. I also tested for interactions between measures of social-emotional skills and gender or race in the reading model. None of these interactions were significant.

Model Fit. This model was only able to account for 23% of the variance between children in the intercept, and 14%, 4%, 0%, and 3% of the between-child variance in the kindergarten, kindergarten to first grade, first to third grade, and third to fifth grade growth rates. Thirty-one percent of the variance across schools in kindergarten was explained by this model. This small amount of explained variance is consistent with the amount of variability explained by others using the ECLS-K data and other similar large-scale survey datasets. One reason for the small amount of explained variability in the reading outcome could be that I have left out key covariates that could predict reading achievement. However, I have included as many measures as are available in the ECLS-K that other researchers have identified as predictive of reading achievement. Another possible reason for the small amount of explained variability in the reading outcome is that some of the variability in the outcome is error, and thus cannot be predicted. I also checked model fit by testing whether the residuals were normal, and found they appeared to be so, based on a visual examination of a graph of the residuals.

Math Achievement Model Results

I now turn to the results from the hierarchical cross-classified model of math achievement. Unstandardized coefficients and standard errors from the math achievement model are presented in Table 5.3. This table only includes measures of health and social-emotional skills and a few key measures of a child's background. Table A.5 in the Appendix contains the full results with all the variables included in the analysis. As with the reading achievement results, I included results from two separate models in Table 5.3, one where regulatory behaviors was included as a stable factor and another where regulatory behaviors was included as a time-varying covariate. At the bottom of Table 5.3 I report how well this model did in predicting variability in math achievement, by listing the proportion of variance in the random effects for the intercept and growth rates

Table 5.3
Unstandardized coefficients and standard errors for covariates of interest from HCM models predicting kindergarten to fifth grade mathematics achievement

	Math IRT Scale Score *with stable regulatory behaviors*		**Math IRT Scale Score** *with time-varying regulatory behaviors*	
	β	*SE*	β	*SE*
Random Level 1 Effects				
π_0: Achievement at Fall Kindergarten				
θ_0: Initial Fall K Achievement	23.247	0.460 **	23.430	0.461**
β_{01}: Average Regulatory Behaviors	2.054	0.091 **		
β_{02}: Birth Weight	0.012	0.003 **	0.013	0.003**
β_{03}: Premature	-0.192	0.246	-0.204	0.250
β_{04}: Average BMI	-0.231	0.081 **	-0.272	0.081**
β_{05}: Food Insecurity	-0.597	0.327 ^	-0.682	0.329*
β_{06}: Male	1.385	0.168 **	0.940	0.166**
β_{07}: Black	-1.489	0.302 **	-1.869	0.303**
β_{08}: Hispanic	-2.473	0.305 **	-2.470	0.308**
β_{09}: Asian	1.842	0.389 **	2.226	0.391**
β_{010}: Other	-1.330	0.386 **	-1.385	0.389**
β_{016}: Average Home Risk Index	-0.932	0.094 **	-1.019	0.094**
π_1: Linear Growth Rate from Fall Kindergarten to Spring Kindergarten				
θ_1: Fall K to Spring K Increase in Achievement	9.792	0.356 **	9.896	0.355**
β_{11}: Average Regulatory Behaviors	1.133	0.071 **		
β_{12}: Birth Weight	-0.002	0.002	-0.002	0.002
β_{13}: Premature	-0.125	0.205	-0.132	0.205
β_{14}: Average BMI	0.080	0.065	0.062	0.065
β_{15}: Food Insecurity	-0.009	0.267	-0.041	0.268
β_{16}: Male	1.113	0.133 **	0.897	0.131**
β_{17}: Black	-1.474	0.268 **	-1.690	0.267**
β_{18}: Hispanic	-0.743	0.251 **	-0.732	0.251**
β_{19}: Asian	0.098	0.347	0.288	0.347
β_{110}: Other	-0.355	0.332	-0.388	0.334
β_{116}: Average Home Risk Index	-0.134	0.073 ^	-0.172	0.073*
π_2: Linear Growth Rate from Spring Kindergarten to Spring Grade 1				
θ_2: Spring K to Spring G1 Increase in Achievement	24.490	0.398 **	24.635	0.400**
β_{21}: Average Regulatory Behaviors	1.317	0.118 **		
β_{22}: Birth Weight	0.003	0.004	0.004	0.004
β_{23}: Premature	0.058	0.309	0.063	0.310
β_{24}: Average BMI	-0.181	0.107 ^	-0.196	0.107^
β_{25}: Food Insecurity	0.175	0.460	0.113	0.463

	Math IRT Scale Score *with stable regulatory behaviors*		Math IRT Scale Score *with time-varying regulatory behaviors*	
	β	*SE*	*β*	*SE*
β_{26}: Male	1.951	0.222 **	1.670	0.219**
β_{27}: Black	-2.563	0.404 **	-2.769	0.405**
β_{28}: Hispanic	-0.487	0.381	-0.501	0.383
β_{29}: Asian	-1.380	0.540 *	-1.288	0.540*
β_{210}: Other	-1.472	0.528 **	-1.500	0.531**
β_{216}: Average Home Risk Index	-0.427	0.121 **	-0.479	0.121**
π_3: Linear Growth Rate from Spring Grade 1 to Spring Grade 3				
θ_3: Spring G1 to Spring G3 Increase in Achievement	32.644	0.493 **	32.780	0.497**
β_{31}: Average Regulatory Behaviors	1.783	0.149 **		
β_{32}: Birth Weight	0.005	0.005	0.005	0.005
β_{33}: Premature	0.517	0.399	0.449	0.401
β_{34}: Average BMI	0.145	0.134	0.141	0.134
β_{35}: Food Insecurity	-1.459	0.550 **	-1.429	0.549**
β_{36}: Male	2.978	0.275 **	2.719	0.273**
β_{37}: Black	-3.757	0.577 **	-3.976	0.581**
β_{38}: Hispanic	-1.325	0.469 **	-1.399	0.468**
β_{39}: Asian	1.260	0.659 ^	1.354	0.657*
β_{310}: Other	-1.259	0.631 *	-1.360	0.635*
β_{316}: Average Home Risk Index	-1.007	0.157 **	-1.076	0.157**
π_4: Linear Growth Rate from Spring Grade 3 to Spring Grade 5				
θ_4: Spring G3 to Spring G5 Increase in Achievement	20.152	0.550 **	20.076	0.550**
β_{41}: Average Regulatory Behaviors	-0.057	0.129		
β_{42}: Birth Weight	0.004	0.004	0.003	0.004
β_{43}: Premature	0.381	0.328	0.391	0.329
β_{44}: Average BMI	0.121	0.119	0.125	0.119
β_{45}: Food Insecurity	0.514	0.475	0.551	0.476
β_{46}: Male	0.426	0.249 ^	0.639	0.249*
β_{47}: Black	0.130	0.572	0.445	0.576
β_{48}: Hispanic	1.182	0.414 **	1.220	0.415**
β_{49}: Asian	1.748	0.700 *	1.540	0.697*
β_{410}: Other	1.519	0.582 **	1.521	0.583**
β_{416}: Average Home Risk Index	0.150	0.134	0.195	0.134
π_5 - π_{35}: Time-Varying Covariates				
θ_5: Regulatory Behaviors			1.576	0.083**
θ_6: Internalizing Problems	-0.373	0.048 **	-0.298	0.050**
θ_7: Health Scale	0.334	0.076 **	0.367	0.076**

	Math IRT Scale Score *with stable regulatory behaviors*		Math IRT Scale Score *with time-varying regulatory behaviors*	
	β	*SE*	*β*	*SE*
θ_8: Disability Status	-1.068	0.222 **	-1.062	0.223**
θ_9: No Doctor/Dentist Visit in past year	0.042	0.137	0.042	0.137
θ_{10}: No Health Insurance	-0.192	0.174	-0.195	0.174
θ_{24}: Regulatory Behaviors* K Growth Rate			0.954	0.070**
θ_{27}: Regulatory Behaviors* K-G1 Growth Rate			-0.767	0.074**
θ_{30}: Regulatory Behaviors* Spring G3 Growth Rate			-0.074	0.061
θ_{33}: Regulatory Behaviors* G3-G5 Growth Rate			-0.355	0.043**
θ_{25}: Health Scale * K Growth Rate	0.052	0.067	0.074	0.067
θ_{28}: Health Scale * K-G1 Growth Rate	0.095	0.108	0.085	0.108
θ_{31}: Health Scale * G1-G3 Growth Rate	-0.188	0.112 ^	-0.201	0.112^
θ_{34}: Health Scale * G3-G5 Growth Rate	-0.003	0.068	-0.018	0.068
θ_{26}: Disability Status * K Growth Rate	-0.461	0.197 *	-0.463	0.198*
θ_{29}: Disability Status * K-G1 Growth Rate	-0.737	0.315 *	-0.773	0.316*
θ_{32}: Disability Status * G1-G3 Growth Rate	1.303	0.326 **	1.237	0.327**
θ_{35}: Disability Status * G3-G5 Growth Rate	0.320	0.205	0.345	0.206^
% Variance in Error and Random Effects Explained by Model				
b_{00}: Fall Kindergarten Intercept (Level 2) Random Child Effect	35.28%		33.99%	
b_{20}: Spring K to Spring Growth Rate Random Child Effect	9.73%		9.33%	
b_{30}: Spring G1 to Spring G3 Growth Rate Random Child Effect	6.76%		6.21%	
b_{40}: Spring G3 to Spring G5 Growth Rate Random Child Effect	0.61%		0.00%	
c_{10}: Fall K to Spring K Growth Rate Random School Effect	35.81%		34.22%	
e: Level 1 Error	3.43%		2.89%	

~ $p < .10$ * $p < .05$ ** $p < .01$

Note: Coefficients in this table are unstandardized.

that have been explained by adding all of my predictors to the model.

The unstandardized coefficients across the two math achievement models in Table 5.3 were fairly consistent. Because of this, I focus my discussion on results from the stable regulatory behaviors model. Similar to the results from the reading achievement models, many of the measures of health and social-emotional skills were statistically significant in the math achievement models. Of the time-varying health measures, children's overall health (θ_7), disability status (θ_8) , and whether or not the child had health insurance (θ_{10}) were all statistically significant, with overall health having a positive effect on math and the other two having a negative effect on math achievement. Whether or not a child visited a doctor or dentist during the past year was not statistically significant. The interaction of disability status with kindergarten growth (θ_{26}) and first to third grade growth (θ_{32}) were statistically significant, and the interaction with the other two growth rates (θ_{29} and θ_{35}) were marginally significant. As was the case with results from the reading achievement models, the effect sizes of the remaining measures of children's health had mixed effects. Of these, birth weight (β_{02}) had a positive effect, and average BMI (β_{04}) and food insecurity (β_{05}) had a statistically significant negative effect on math achievement at kindergarten entry. Average BMI also had a negative effect on the math growth rate from kindergarten to first grade. Food insecurity had a statistically significant negative effect on math growth rate from first to third grade. None of the other health measures had a significant effect on math achievement growth rates. Both internalizing problems and regulatory behaviors had a statistically significant effect on math achievement. The internalizing problems had a negative effect on math achievement, while the effect of regulatory behaviors (average or time-varying) on achievement were positive.

As with the results from the reading model, the unstandardized coefficients in the math model were not easily comparable across predictors in the model or over time. I thus transformed these coefficients into standardized effect sizes. In Figure 5.3, I graphed the standardized effect sizes of measures of health and social-emotional skills that had a statistically significant effect on math achievement over time. Each line on the graph represents the effect size for a single measure, holding all other covariates in the model constant. Table 5.4 presents the effect size at each point in time in two ways: (1) the effect in points on the reading IRT scale score, and (2) the standardized effect size. In this table, I include only those health and social-emotional skills measures that had a statistically significant effect on math achievement over time. I also included, for

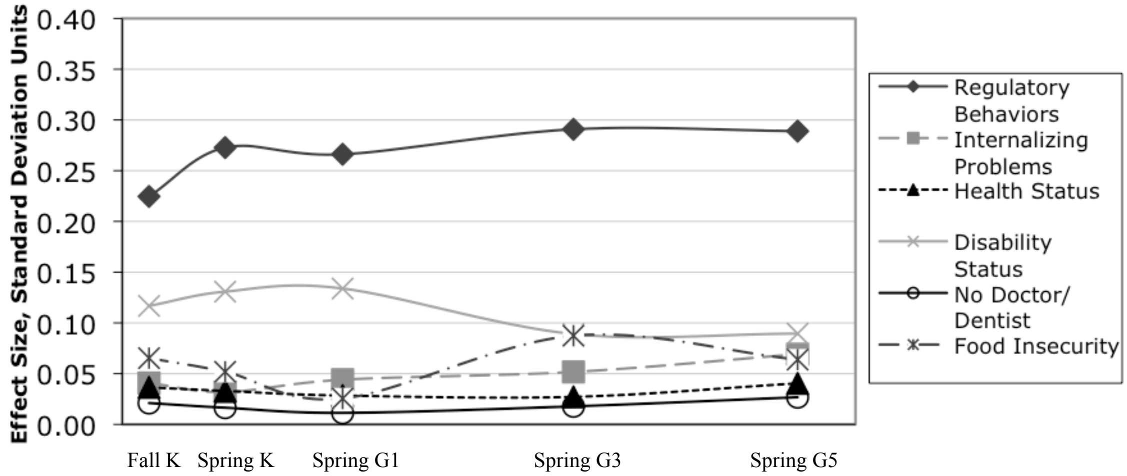
Table 5.4
Effect sizes over time of health and social-emotional skills measures and other key child characteristics on math achievement

		Fall K	Spring K	Spring G1	Spring G3	Spring G5
Average Regulatory Behaviors	*Points**	2.054	3.188	4.505	6.288	6.231
	*SD^*	0.225	0.273	0.266	0.291	0.289
Regulatory Behaviors, Time-Varying	*Points*	1.576	2.531	3.528	5.068	5.336
	SD	0.172	0.217	0.208	0.234	0.247
Internalizing Problems	*Points*	-0.373	-0.373	-0.746	-1.119	-1.492
	SD	-0.041	-0.032	-0.044	-0.052	-0.069
Overall Health Scale	*Points*	0.334	0.386	0.481	0.587	0.872
	SD	0.036	0.033	0.028	0.027	0.040
Disability Status	*Points*	-1.068	-1.529	-2.266	-1.926	-1.929
	SD	-0.117	-0.131	-0.134	-0.089	-0.089
No Health Insurance	*Points*	-0.231	-0.151	-0.332	-0.187	-0.065
	SD	-0.025	-0.013	-0.020	-0.009	-0.003
Food Insecurity	*Points*	-0.597	-0.606	-0.431	-1.890	-1.376
	SD	-0.065	-0.052	-0.025	-0.087	-0.064
Home Risk Index	*Points*	-0.932	-1.066	-1.493	-2.501	-2.351
	SD	-0.102	-0.091	-0.088	-0.116	-0.109
Male	*Points*	1.385	2.498	4.448	7.426	7.852
	SD	0.151	0.214	0.263	0.343	0.364
Black	*Points*	-1.489	-2.963	-5.527	-9.284	-9.154
	SD	-0.163	-0.254	-0.327	-0.429	-0.425

* Points at each time point have been accumulated by adding coefficients from the intercept and all growth rates prior to and including a given time point.

^ Standard deviation effect sizes calculated by dividing total points at each time point by the raw standard deviation of math achievement for that time point.

Figure 5.3
The standardized effect size of health and social-emotional skills measures on math achievement from kindergarten through fifth grade



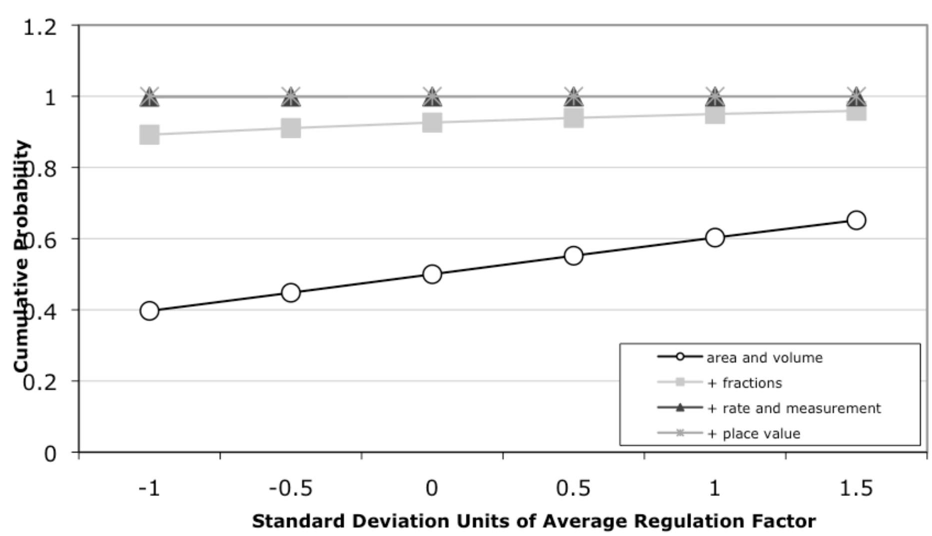
comparison purposes, the effect size of three measures that previous research has shown to have substantial effects on reading achievement: gender, black race/ethnicity, and average home risk index.

Social-Emotional Skills. I found, as seen in Table 5.4, that a child's regulatory behaviors had a larger effect size (of 0.225 sd) on math achievement at kindergarten entry than all other health and social-emotional skills measures. The regulatory behaviors effect size on initial kindergarten math achievement was also larger than the effect of being black (-0.163 sd), male (0.151 sd), or having a bad home risk index (-0.102 sd). Tracking the growth in the effect size of regulatory behaviors on math achievement over time, as shown in Figure 5.3, I found the effect size on math increased through the end of kindergarten, was then basically maintained through fifth grade, and ended with an effect of 0.289 standard deviations. The size of the effect of the time-varying regulatory behaviors factor on math achievement was slightly smaller than that of the stable factor, with an effect size of 0.172 standard deviations at kindergarten entry and 0.247 sd by the end of fifth grade.

Comparing the size of the effect of regulatory behaviors on math achievement over time to that of other measures, I found that none of the other health and social-emotional skills measures came close to having the same magnitude of effect. By the end of first grade, the effect of regulatory behaviors on achievement was similar to that of being black and male, but then the effect size of being male and black on math achievement outgrew that of regulatory behaviors, though not by large amounts. Being male and black had effect sizes of 0.364 and -0.425 on fifth grade math achievement, respectively, compared to the effect size of 0.289 for regulatory behaviors.

In order to understand better what this math gap due to regulatory behaviors means in terms of what has been learned, I report the results from a 2-level HLM ordinal model with fifth grade math proficiencies as the outcome. The cumulative probability of being proficient on the highest math proficiency levels is presented graphically in Figure 5.4. Children who were one standard deviation below the mean on regulatory behaviors had a 40% probability of being proficient on the highest proficiency level (area and volume), compared to 60% of children who were one standard deviation above the mean. The probability of math proficiency on either of the two highest proficiency levels (area and volume and fractions) was 89% for children who were one standard deviation below the mean and 95% for those one standard deviation above the mean of regulatory behaviors. The cumulative probability for all the other math proficiencies were similar across levels of regulatory behaviors.

Figure 5.4
Predicted cumulative probability of fifth grade math proficiency by average regulatory behaviors



Turning to the other measures of social-emotional skills, I found that the effect size of internalizing problems on math achievement was much smaller than that of regulatory behaviors at each time point, as seen in Figure 5.3. The initial kindergarten effect size of internalizing problems on math achievement was -0.041 sd. This was basically maintained through fifth grade, ending with an effect size of -0.069 sd.

Health. Whether or not a child had any disability, the child's food insecurity status had the largest effect on mathematics achievement over time of all the health measures. The math gap between children with and without any disability was -0.117 sd at kindergarten entry. This gap increased slightly through kindergarten, then dropped by third grade, ending in fifth grade slightly smaller than the initial kindergarten gap, at -0.089 sd. The initial and the final math gap due to food insecurity were similar to each other, though the size of the gap in the intermittent years fluctuated up and down. The effect sizes of the remaining health measures on math achievement were all small, with minor fluctuations over time. The largest increase in the effect size for the health factors on math achievement occurred at kindergarten entry, with minimal changes in the size of the effect in subsequent grades.

Other Risk Factors. As I did with the reading model, I checked for the possibility of health and social-emotional skills measures acting as mediators for gender, race, and other background characteristics. I found that gender, measures of race, and the home/family risk index were partially mediated by the addition of health and social-emotional skills measures to the math model. I also tested to see if average regulatory behaviors interacted with gender or race. I found a positive interaction for gender and average regulatory behaviors on the kindergarten, kindergarten to first grade, and first to third grade growth rates. The interaction effect, though statistically significant, was small. For example, the unstandardized coefficient on the kindergarten growth rate was .4, which correlated to a .05 sd effect size. This interaction effect indicated that the effect of regulatory behaviors on math achievement was stronger for boys than for girls. Boys one standard deviation above the mean on regulatory behaviors had higher math achievement than girls, while boys one standard deviation below the mean on regulatory behaviors had lower math achievement scores than girls with the same level of regulatory behaviors.

Model Fit. This model was only able to account for 35% of the variance between children in the intercept, and 10%, 7%, 1%, and 3% of the between-child variance in the kindergarten, kindergarten to first grade, first to third grade, and third to fifth grade growth rates. Thirty-five percent of the between-school variance in kindergarten was explained by this model. As mentioned earlier, this small amount of explained variance is consistent with the amount of variability explained by others using the ECLS-K data and other similar large-scale survey datasets. Reasons for the small variability could be predictors left out of the model or the amount of error in the outcome.

Sensitivity of Math and Reading Results to Specification Error

One of my concerns with both the reading and math achievement models presented above was that the results could be sensitive to the types of measures included as well as potential confounding variables. In an attempt to determine how sensitive my results were to these issues, I ran two additional models—the first to determine if the growing gaps found in the IRT scale score were actually just a result of the increasing variability in scale scores over time, and the second to test if any unmeasured confounding variables could dramatically change my results. To address the scaling issue of the outcome, I re-ran both the reading and math HCM models with standardized scores as the dependent variable. I tested the sensitivity of the original models to confounding variable bias by running a new model using propensity score stratification.

In Figures 5.5 and 5.6 I present the standardized effect size of average regulatory behaviors over time on reading and math achievement, controlling for all other covariates, from the original models alongside the effect sizes found from the two new analyses. The first bar at each time point represents the original models with scale scores as the outcome, the second bar represents a model with standardized scores as the outcome, and the final bar at each time point represents a scale score model with propensity score strata included as covariates. The whiskers in this plot represented the standard errors (±2 SE), allowing me to determine whether the differences in effect sizes across models were actually different from each other. While I only present the results here for similarities and differences of the regulatory behaviors effect on reading and math achievement, I also checked the differences on the other measures of interest and

Figure 5.6
Effect of average regulatory behaviors on math achievement from two statistical models with different dependent measures of achievement and a third model using propensity strata

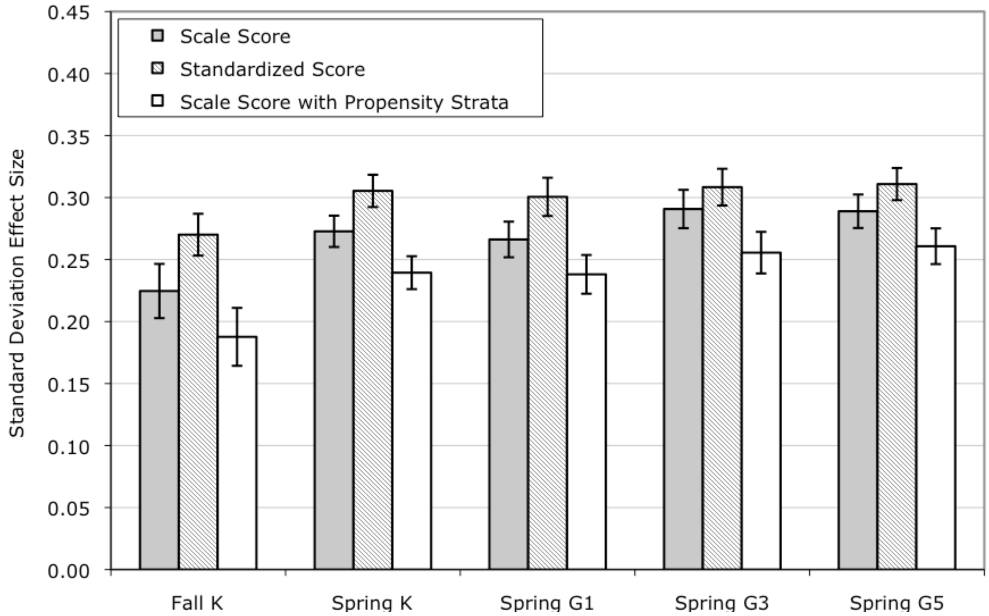
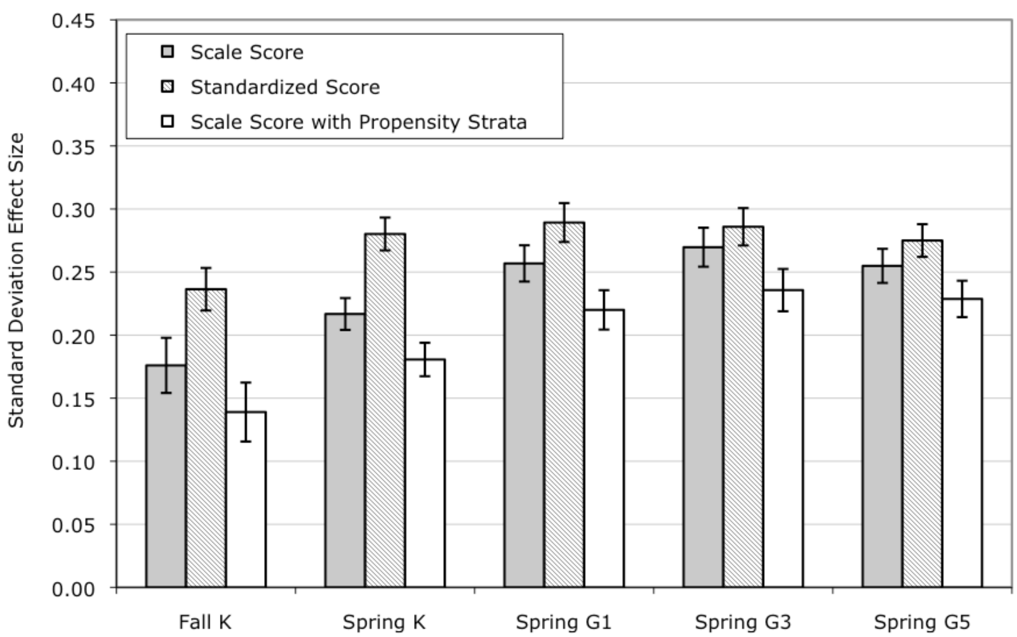


Figure 5.5
Effect of average regulatory behaviors on reading achievement from two statistical models with different dependent measures of achievement and a third model using propensity strata



found similar results to what I present below.

First, with regards to the scaling of the outcome, I found the size and direction of the effect of regulatory behaviors on standardized reading and math scores over time to follow a similar pattern as those of the IRT scale score, with only a few discrepancies. The effect size for regulatory behaviors on the standardized reading score began higher than for that of the IRT scale score, and only increased slightly over kindergarten before leveling off through fifth grade. By first grade, the differences between the scale score and standardized score were not statistically significant. Thinking of the standardized scores as a blunter longitudinal indicator, it is not unexpected that I only found an increase in the effect size over kindergarten on this measure, as the largest changes in the effect size of regulatory behaviors on the reading scale score occurred in kindergarten, with increases in the effect size in subsequent years reflecting much smaller change. Comparing the effect size of regulatory behaviors on standardized scores to scale scores in the math model, as seen in Figure 5.6, I found the effect sizes were similar at each time point. This led me to believe that results from the scale score over time are valid.

I also ran a reading and math scale score model with propensity score strata to address the possibility of confounding variables substantially changing my results. I found that the size of the effect of regulatory behaviors on both math and reading achievement were significantly lower at each time point than the original scale score model. The drop in the size of the effect, however, was not large, and the effect size could still be considered substantial, leading me to believe that confounding variables were not a serious threat to the validity of the results presented above.

Research Question 2: Exploring the Combined Effect of Multiple Risks on Reading and Math Achievement

All of the results presented up to this point focused on the effect of individual risk factors on math and reading achievement, holding all other measures in the model constant. While these results were interesting and helpful in understanding what had the largest effect on achievement, they did not provide a comprehensive picture of the magnitude of the academic achievement gap for children at risk on multiple factors. Past research, in fact, has shown that risk factors do tend to cluster in individuals (Gutman et al., 2003). In this section I first explored what other factors tended to cluster with the risk of regulatory behaviors problems[9]. To do this, I created three risk groups based on the regulatory behaviors factor. Those who were one standard deviation below the average or worse on the regulatory behaviors factor were labeled as the high-risk group (16.6% of sample), those one standard deviation above the mean or better were labeled as the low-risk group (16.2%), and the remaining were labeled as the average risk group (67.2%). I calculated the means and standard deviations of a wide range of risk factors for each group to see if differences existed across groups. I next turned my attention to determining the size of the total achievement gap for children in the high-risk group compared to those in the average and low-risk groups. To calculate the total gap, I plugged the risk-group averages for each of the covariates in my model (both health and social-emotional skills measures as well as all other child, home, teacher, and school measures) into the reading and math achievement models to obtain model-based estimates of the size of the achievement gap between risk groups. Results are presented below, first of the covariate means by regulatory behaviors risk group, and next of the achievement gap by regulatory behaviors risk group.

Means for a wide range of child, family, teacher, and school measures by regulatory behaviors risk group are presented in Table 5.5. The differences across risk groups in the means of almost all of these measures were statistically significant. Using this table, I was able to create a profile of the type of child who had regulatory behaviors problems and determine the other factors they were at risk for having. I found that children in the high risk group for regulatory behaviors problems were at greater risk for having poorer overall health and having a physical or mental health disability than children in the average and low-risk groups. They were also less likely to have visited a doctor or dentist within the last year or to have health insurance. The majority of high-risk children were male, and significantly more racially black children and children came from families and homes with a higher risk index. Children in the high-risk group were in

[9] I grouped children by regulation rather than by any of the other health and social-emotional skill measures of interest, because regulation consistently had the largest effect size on both reading and math achievement of all my measures of interest, and I was interested in determining what proportion of the full achievement gap was due to differences in regulation.

Table 5.5
Means and standard deviations of kindergarten child, family, teacher and school risk factors by regulatory behaviors risk group

	High Risk	Average Risk	Low Risk	
Social-Emotional Skills Measures				
Regulatory Behaviors	-1.668 (0.534)	0.100 (0.552)	1.294 (0.210)	**
Internalizing Problems	0.399 (1.122)	-0.018 (0.977)	-0.331 (0.788)	**
Externalizing Problems	1.282 (1.117)	-0.138 (0.766)	-0.745 (0.432)	**
Interpersonal Skills	-0.966 (0.838)	0.020 (0.880)	0.886 (0.617)	**
Self-Control	-1.117 (0.858)	0.061 (0.844)	0.886 (0.617)	**
Approaches to Learning	-0.996 (0.846)	0.020 (0.887)	0.933 (0.570)	**
Child Health Measures				
Health Scale	-0.170 (1.078)	0.028 (0.980)	0.179 (0.892)	**
Disability Status	18.7%	11.4%	9.8%	**
No Doctor/Dentist Visit in past year	22.8%	19.5%	16.2%	**
No Health Insurance	11.3%	9.2%	5.4%	**
Food Insecurity	12.2%	8.3%	4.9%	**
BMI	0.134 (1.070)	0.001 (1.000)	-0.140 (0.902)	**
Birth Weight	117.098 (31.967)	118.770 (31.967)	118.860 (28.001)	~
Premature	17.4%	16.3%	15.8%	
Child/Family Background Measures				
Male	71.7%	50.7%	28.8%	**
Black	22.2%	10.3%	5.1%	**
Hispanic	19.5%	19.7%	14.3%	**
Asian	3.3%	7.2%	9.4%	**
Other	7.5%	5.3%	3.8%	**
Age	78.313 (4.970)	78.345 (4.366)	78.977 (4.081)	**
No English in Home	13.4%	17.2%	13.1%	**
Home Risk Index	0.214 (0.924)	-0.001 (0.850)	-0.213 (0.728)	**

	High Risk	Average Risk	Low Risk	
Classroom/Teacher Measures				
Full Day Kindergarten	62.5%	54.3%	50.8%	**
Class Size	23.235 (8.528)	23.142 (8.654)	23.207 (8.445)	
% Minority in Class	45.7%	39.0%	31.4%	**
% Disability in class	8.2%	7.4%	7.0%	**
% Read below grade level in class	18.0%	15.9%	13.6%	**
Teacher rating of Class Behavior	0.329 (0.829)	0.559 (0.823)	0.711 (0.847)	**
Parent volunteered in class	36.7%	53.6%	69.3%	**
# volunteer hours	3.050 (5.021)	3.479 (5.296)	3.778 (5.700)	**
School Measures				
Northeast	15.1%	18.7%	21.2%	**
South	36.7%	31.1%	31.3%	**
West	21.3%	24.2%	23.1%	**
Private School	15.3%	31.8%	33.5%	**
% Students reading at or above grade level	57.9%	62.6%	65.9%	**
% Students math at or above grade level	59.3%	63.6%	66.7%	**
Special Ed FTE	2.941 (3.824)	2.543 (2.572)	2.579 (2.564)	**
Nurse FTE	0.492 (0.631)	0.482 (0.786)	0.513 (1.209)	
Policy for Quiet/Orderly Environment	41.9%	40.0%	36.8%	**
School Neighborhood Safety	-0.067 (0.997)	0.0634 (0.945)	0.267 (0.867)	**

~ $p < .10$ * $p < .05$ ** $p < .01$

classrooms, on-average, that provided more full-day kindergarten and had poorer behaving classrooms, a greater percentage of students reading below grade level, and fewer parents volunteering in the classroom. High-risk children were less likely to attend private school, and more likely to live in the South and in less safe neighborhoods than children in the other risk groups.

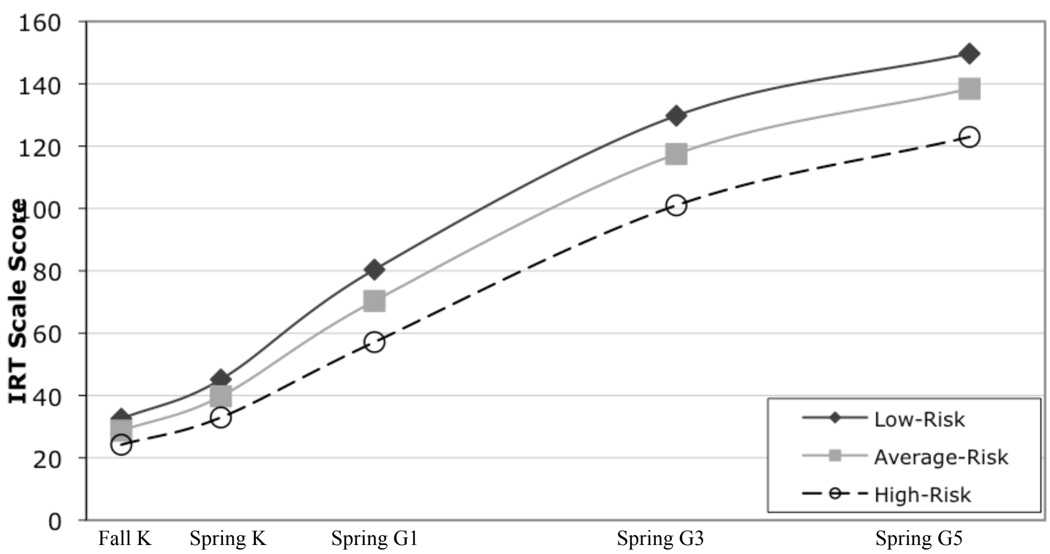
I next present the results of the total achievement gaps by regulatory behaviors risk group. Model-based estimates of the reading achievement gap are presented in Table

5.6. I included the gap between high-risk and average-risk children, as well as the gap between high-risk and low-risk children, in this table. The reading achievement gaps were substantial. Children in the high-risk group entered school 0.43 standard deviations below their average risk peers, and 0.80 standard deviations below their low-risk peers. This gap grows through third grade and is maintained through fifth grade, resulting in a 0.66 sd gap with the average risk group and a 1.14 sd gap with the low-risk group. Figure 5.7 illustrates the magnitude of the growing gap over time due to the diverging

Table 5.6
Reading achievement gaps between regulatory behaviors risk groups

	High-Risk vs. Average-Risk Gaps		**High-Risk vs. Low-Risk Gaps**	
	Points	*Effect Size*	*Points*	*Effect Size*
Beg. K	4.51	0.43	8.33	0.80
End K	6.81	0.49	12.21	0.88
End G1	13.21	0.59	23.25	1.04
End G3	16.46	0.65	28.76	1.13
End G5	15.46	0.66	26.75	1.14

Figure 5.7
Average reading achievement trajectories
by regulatory behaviors risk group



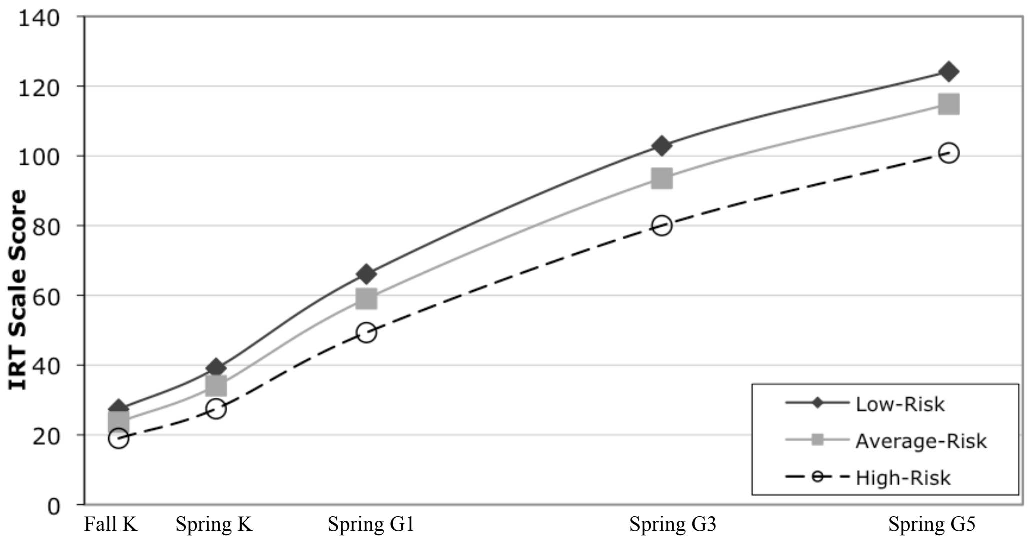
trajectories of children in the different regulatory behaviors risk groups. This figure charts the IRT scale score point differences by risk group.

The math achievement gaps by risk group are presented in Table 5.7. Once again, I presented the math gaps between high-risk and average-risk children, as well as the gaps between high-risk and low-risk children in this table. The math achievement gaps, as seen in Table 5.7, were slightly larger than the reading achievement gaps. Children in the high-risk group entered school already 0.51 standard deviations behind their average-risk peers, and 0.91 sd behind their low-risk peers. This gap grew slowly through each grade, ending fifth grade with a 0.65 sd gap in math achievement between the high risk and average risk groups and 1.08 sd between the high risk and low risk groups. Figure 5.8 illustrates the magnitude of the growing math gap over time, due to the diverging trajectories of children in the different regulatory behaviors risk groups.

Table 5.7
Mach achievement gaps between regulatory behaviors risk groups

	High-Risk vs. Average-Risk Gaps		**High-Risk vs. Low-Risk Gaps**	
	Points	*Effect Size*	*Points*	*Effect Size*
Beg. K	4.64	0.51	8.35	0.91
End K	6.60	0.56	11.62	0.99
End G1	9.71	0.57	16.78	0.99
End G3	13.53	0.63	22.87	1.06
End G5	14.01	0.65	23.33	1.08

Figure 5.8
Average math achievement trajectories by regulatory behaviors risk group



Chapter 6
Discussion and Conclusions

The purposes of this dissertation are two-fold. First, to provide a comprehensive review of the constructs and measures comprising the domains of health and social-emotional skills and the knowledge regarding their association with academic achievement. Second, to determine how health and social-emotional skills relate to reading and math achievement.

Regarding the first purpose, my literature review shows that health and social-emotional skills constructs can be grouped into four general categories: (1) physical health, encompassing areas such as chronic illnesses, physical disabilities, weight, nutrition, health at birth, and general health status; (2) problem behaviors or mental health problems, such as externalizing and internalizing problems; (3) social competence or mental health, characterized by skills such as interacting positively with peers, communicating effectively, sharing, cooperating, and helping others; and (4) learning-related social skills, including self-control, listening, working in groups, and following instructions. These last three categories comprise the domain of social-emotional skills. I found in the literature that social-emotional skills were closely related to self-regulation. Self-regulation skills, such as attention, memory, and inhibition, were believed by many scholars to form the basis for more overt social-emotional skills, though the actual causal pathways have not been firmly identified.

With an understanding of what comprised the domains of health and social-emotional skills, I next reviewed the literature that dealt with the relationship between these domains and academic achievement with the aim of identifying aspects of health and social-emotional skills that strongly relate to achievement. While there is substantial evidence of a correlation between both health and social-emotional skills and to academic achievement, relatively few studies have addressed the three areas of social-emotional skills in conjunction with each other (e.g., Claessens, Duncan & Engel, 2008; Duncan et al., 2007; Howse, Calkins et al., 2003; Spira & Fischel, 2005). Most of these studies found that attention and other learning-related and regulatory skills had the strongest effect on achievement, and that after controlling for learning-related skills, measures of social skills and problem behaviors were no longer significant in their models of achievement.

One issue that has been largely ignored in the literature is that both health and social-emotional skills can change over time. Accounting for the variability of these factors over time should lead to more accurate estimates of effect sizes that will estimates based on a measurement from a single time point. However, I found only one study that accounted for repeated measurements over time in their longitudinal achievement models (Gutman et al., 2003).

The second purpose of my dissertation was to perform a longitudinal analysis of data from the ECLS-K to explore the relationship between health and social-emotional skills and reading and math achievement over time. With this analysis, I addressed two shortcomings I discovered in the current literature. First, by including multiple measures of health and social-emotional skills in my model, along with a large set of commonly used controls, I determined which measures within these domains had the strongest effect on reading and math achievement. Second, I accounted for the time-varying nature of health and social-emotional skills. In doing this analysis I assume that by accounting for the variability of these factors over time, through the use of repeated measures, I will achieve better estimates of how health and social-emotional skills affect achievement over time.

Using hierarchical, cross-classified models of reading and math achievement from kindergarten to fifth grade, I found that many of the health and social-emotional skills measures I included had statistically significant effects on reading and math achievement at kindergarten entry and throughout elementary school, holding all other measures in the model constant. In particular, a child's regulatory behaviors had, by far, the strongest effect of all health and social-emotional skills measures. I found that taking into account the time-varying nature of these factors led to some effects being maintained over time (overall health, interpersonal skills, and whether or not a child has health insurance), while other factors had effect sizes that increased over time (regulatory behaviors and disability status).

In the following sections, I elaborate on these findings. I begin by discussing the key results stemming from the social-emotional skills measures, followed by a discussion of results from the health measures. I then discuss the findings of the gap analysis for children from different risk groups. I conclude this chapter by reviewing the strengths and limitations of this study and suggested directions for future research.

Social-emotional Skills

In this dissertation I used the term "social-emotional skills" to refer to the set of skills and behaviors encompassing social skills, problem behaviors, and learning-related social skills. Previous research has shown that all of these areas affect achievement, though when accounting for all three areas simultaneously in statistical models, the literature suggested that only measures of learning-related skills, such as attention, have an independent effect on achievement. I initially wanted to try to duplicate that finding with my analysis of the ECLS-K data. However, although the ECLS-K did have teachers rate children on all three of these areas using the Social Rating Scale (SRS), I found that four of the five scales were highly correlated and including them all in the same model could lead to problems with multicollinearity. The four correlated scales were self-control, approaches to learning (e.g. attention), interpersonal skills, and externalizing problems. I chose to combine these four scales into a single factor, which I termed regulatory behaviors, based on the assumption that teachers' ratings of children's social skills may be tapping into the underlying latent construct of self-regulation, while also recognizing that the factor of combined SRS scales is based on teacher perceptions of children's behavior. The decision to label this factor regulatory behaviors was also an attempt to distinguish it from more current measures of children's regulation through direct assessment. While there is evidence that direct measurements and observations are highly correlated (e.g. Ponitz, et. al., 2007), there is still uncertainty as to what they are each truly measuring, which could lead to different interpretations of results and conclusions based on results.

Although combining four of the five SRS scales limited my ability to determine which of the three areas of social-emotional skills had the strongest independent effect on achievement, as I was left with only internalizing problems and regulatory behaviors, I did still find some very interesting results based on the regulatory behaviors factor, which I discuss in more detail below.

Stability of Measures Over Time

During data preparation, I examined how the regulatory behaviors factor behaved over time in order to inform how it should be included in my longitudinal models of reading and math achievement. Using a simple measurement model, I found that a majority of the variability in regulatory behaviors (67%) was due to differences between children, leaving only 33% of the variability in regulatory behaviors to differences over time[10]. I was, admittedly, surprised by this result, as most child development literature suggested that middle childhood is a time of considerable growth and change in a child's social-emotional skills and behaviors (Huston & Ripke, 2006; Moffit et al., 2002; Shonkoff & Phillips, 2000).

I proceeded to run a measurement model on the original SRS scales that comprised the regulatory behaviors factor to see if they were also somewhat stable over time. I found that between 45%-60% of the variability in the original SRS scales was due to differences over time, indicating that these were more time-variable than the regulatory behaviors scale. One of the reasons for this could be the measurement model and estimation methods I used. Because I did not have the original items used to create the SRS scales, I had to use a simplified measurement model and fairly crude technique to estimate the percent of variability between children over time (a method I did not need to use for the regulatory behaviors factor) which could provide incorrect estimates of variability over time.

If, however, the results from the measurement models were valid, and the individual scales really were more variable over time than the latent factor of regulatory behaviors, this result could come about for number of different reasons. First, changes

[10] As explained in Chapter 4, the estimate of between-child variability in the outcome was computed after removing the error term from the model. This error term included the systematic error due to different raters over time.

over time could be due to age-related maturation of social behaviors. Although if this were the case, I would expect to see similar age-related maturation and growth in the regulatory behaviors scale. A second reason for the variability of the individual behavior scales and relative stability of regulatory behaviors could be that children's social behaviors fluctuate over time, while their underlying regulatory skills remain fairly stable. Evidence from researchers studying children's social competence was able to help shed light onto why this could be. Competence has been described as the upper limit of a child's ability and it is highly dependent on the context where the child is functioning (Fischer et al, 1993; Masten & Coatsworth, 1998). For example, while children may have acquired a specific level of social skills, they might not demonstrate competence at that skill level, based on their observable behaviors in a specific context, due to the challenges, stresses, or lack of scaffolding in their surrounding environment. Alternatively, certain types of contexts could provide more structure and support, enabling children to perform at higher skill levels, thus demonstrating competence they might not have displayed otherwise (Fischer et al, 1993; Masten & Coatsworth, 1998; Vygotsky, 1978; Welsh & Bierman, 2001). As children in the ECLS-K moved from grade to grade, they experienced a sequence of potentially very different environments. Thus, I would expect to see some fluctuations in their outward social behaviors, even when their underlying skill level did not change. In other words, I might expect to see fluctuations in teacher ratings of observed behaviors (e.g., the SRS items and scales) while the child's latent, underlying skills remained fairly consistent over time (e.g., regulatory behaviors).

From an educator's standpoint, I would hope that a child's self-regulation skills and behaviors could be changed, or more specifically, improved over time. Interventions such as the Tools of the Mind preschool curriculum, in fact, have found that self-regulation and executive function skills can be improved through training and practice (e.g. Barnett et al., 2008). While these skills can be taught when specifically targeted, studies looking at whether normal schooling practices improve self-regulation found no evidence of an effect (Burrage et al., 2008; McCrea, Mueller, & Parrila, 1999; Ponitz, Rimm-Kaufman, Brock & Nathanson, 2009). As the ECLS-K is a nationally-representative sample of children in normal school settings, I would assume that the majority of children in the ECLS-K experienced fairly standard classroom practices – practices that on their own might not lead to improvements in children's self-regulation. In conclusion, I would argue that the amount of stability I found in the regulation scale over time indicated that children's underlying regulation skills were fairly stable, even though the observed behaviors stemming from these skills fluctuated over time. I would add that this conclusion does not mean regulatory skills cannot be changed or improved, but rather that improving social-emotional and regulatory skills might require specific training and practice through targeted interventions.

Effect of Regulatory Behaviors on Reading and Math Achievement

To test the effect of regulatory behaviors on math and reading achievement I ran two separate math and reading achievement models using ECLS-K data, one with regulatory behaviors as a stable factor and one with regulatory behaviors as a time-varying factor. I found that the time-varying and stable specifications of regulatory behaviors produced similar results, with the effect sizes from the stable factor on math and reading achievement being slightly larger. Because of the similar results, I focus my discussion on results from models using the average regulatory behaviors scale over time.

In both the reading and math achievement models, the regulatory behaviors factor had the strongest effect of all the health and social-emotional skills measures. In the reading models, the effect of regulatory behaviors on reading began in kindergarten with an effect size of 0.176, which grew through first grade, and then was basically maintained through fifth grade, ending with an effect size of 0.255. The effect of regulatory behaviors on math achievement started in kindergarten larger than that of reading, with an effect size of 0.225. This effect size grew, mostly over kindergarten and from first through third grade, until fifth grade where an effect size of 0.289 on math achievement was observed. While all of these effect sizes appear to be in the same range, the standard errors are all fairly small (typically around .01), indicating that the differences between effect sizes on reading and math are real, even if they are substantially small. One of the reasons for the stronger effect of regulatory behaviors on math achievement in kindergarten compared to reading could be that more time is spent on reading in most classrooms, providing students with regulatory behaviors problems along with more opportunities to learn, despite regulatory problems.

A number of reasons why poor regulatory behaviors could lead to poor academic achievement in either reading or math exist. Poor regulatory behaviors are demonstrated through poor social skills and problem behaviors, and they also limit a child's ability to navigate and effectively manage the classroom environment and take advantage of learning opportunities (Graziano et al., 2007; Howse et al., 2003; Morgan et al., 2008; Rothbart & Bates, 1998). Regulation skills are required to pay attention, shift attention between tasks, wait, remember and follow directions, compare and contrast ideas, make and follow through on plans, and inhibit inappropriate responses while activating appropriate responses (Dowsett & Livesey, 2000; Gathercole & Pickering, 2000; Kail, 2003; McClelland & Morrison, 2003; Morrison et al., 2009; Rimm-Kaufman et al., 2000; Rueda et al., 2005; Wentzel, 1993; Zelazo et al., 2003). It is not surprising, then, that the effect of regulatory behaviors is so substantial.

Comparing the size of the regulatory behaviors factor to all of the measures in the reading and math models, I found that the only measures with similar or stronger effects on achievement than regulatory behaviors over time were the home risk index, the child's race, and whether or not English was spoken in the home. All other measures describing the child's background, as well as measures of the home, classroom, and school context had smaller effect sizes than did regulatory behaviors. Holding all other measures in the model constant, comparing the effect size of regulatory behaviors to the total achievement gap between children in the high-risk group and the average-risk group (based on estimates using group averages on all risk factors) was quite telling of the importance of this factor. At kindergarten entry, for example, the math achievement gap between the high- and average-risk group was 0.51 sd. The math achievement gap for children with a one standard deviation difference in regulatory behaviors at kindergarten entry was 0.225 sd – almost half the size of the total gap between risk groups. This indicated that for children with regulatory problems, these problems were the reason for almost half of their total math achievement gap at kindergarten entry. A similar proportion of the total achievement gap was due to regulatory behaviors across all grades in both math and reading achievement.

I also compared the size of the effect of regulatory behaviors from my models to those reported in the literature for similar constructs. I found that my effect sizes were generally larger than those previously reported. For example, one study based on an analysis of six longitudinal datasets, including the ECLS-K, found that kindergarten attention (or, more accurately, behaviors indicative of attention), after controlling for other social and behavior measures, had a consistent .1 sd effect size on third or fifth grade reading achievement across datasets (Duncan et al., 2007). There were, of course, a number of differences between my analysis and that of Duncan and colleagues. First, they included all the individual measures of behavior (including attention) in their models, whereas I combined the measures into a single factor. Secondly, their analysis looked at the long-term effects of kindergarten skills on fifth grade achievement, while I incorporated repeated measures over time in my analysis. Despite these differences, I would still argue that while it appears that attention is important, the stable latent factor of regulatory behaviors underlying this learning-related behavior and incorporating a variety of other learning-related skills, has an even larger effect on achievement.

Finally, looking at the pattern of the size of the effect of regulatory behaviors on achievement over time, I found that children entered school already having a substantial gap in reading and math achievement due to regulatory behaviors. This gap then increased most over kindergarten and first grade, before leveling off through the later elementary school grades. This pattern indicated that the key time to intervene on regulatory behaviors, in order to improve academic achievement, would begin in preschool and continue at least through the first years of elementary school, if not longer.

Limitations of the Regulatory Behaviors Results

A number of limitations arise from my use of the regulatory behaviors factor in my models of achievement. First, it is important to remember that my regulatory behaviors factor is not a direct measure of children's regulatory behaviors, but rather, an estimation based on combining four SRS scales. This new factor is a broad, all-inclusive measure, leading to the question of what exactly was represented by this factor. While the high correlations of the four SRS scales indicated some underlying construct tying them together, I cannot know for certain what that latent construct was. I have made the assumption that this latent construct was self-regulation, based on evidence from previous research that suggested self-regulation might underlie more overt social-emotional skills. However, it is possible that this factor was estimating some other underlying construct, such as general good behavior. The results from the regulatory behaviors factor presented in this work should thus be verified by using datasets with more precise, direct measures of regulation.

The broadness of the regulatory behaviors factor could also lead one to question whether the large effect sizes I found were due to my having cast a wider net. While this is a possibility, if my assumption that self-regulation underlies the individual SRS scales is correct, than knowing that the broad area of regulatory behaviors is strongly associated with achievement is valuable. This broad factor has provided an overall estimate of the importance of regulatory behaviors, thus providing a starting point where researchers can continue to investigate more thoroughly, using more precise measures.

Effect of Internalizing Problems on Reading and Math Achievement

While I have focused my discussion thus far on results based on the regulatory behaviors factor, I also included one other measure of social-emotional skills in my reading and math achievement models: internalizing problems. I found that internalizing problems had a statistically significant, though small, effect over time on both reading and math achievement (of about 0.05 sd), after controlling for the other measures in the model. While this might look like a weak, unimportant gap, it was similar in size to the effects of many of the statistically significant measures from the classroom and school ßincluded in my model.

Internalizing problems occur when children inhibit behavior too much. Internalizing problems are harder to identify than externalizing problems, because there are fewer outwardly observable characteristics. Students with internalizing problems are more likely to be quiet and withdrawn, less likely to participate in group learning activities, often receive less attention and instruction from their teacher and have lower motivation to learn (Rapport et al., 2001). Because internalizing problems are difficult to identify, I would expect that the effect sizes found in both reading and math would be lower-bounds of the true effect size.

This result was different from what others have found when including internalizing problems along with other measures of behavior (e.g., Claessens et al., 2008; McLeod & Kaiser, 2004). Claessens and colleagues (2008), for example, found that kindergarten measures of internalizing problems were no longer significant on reading or math achievement in fifth grade, after controlling for other social skills and learning-related behaviors. One reason for the difference between findings could be that I included internalizing problem behaviors as a time-varying covariate, whereas Claessens and colleagues were looking at the long-term effects of kindergarten measures. The fact remains, however, that even though I found a significant effect, it was quite small, leading me to believe that educators interested in improving the academic achievement of their students, should spend the time, effort, and expense of developing an intervention to more effectively target regulatory behaviors.

Physical Health

Along with social-emotional skills, I was also interested in the effect of measures of health on math and reading achievement. I found from my analysis of ECLS-K data a number of statistically significant health measures in both the reading and math models, including disability status, overall health status, food insecurity, and whether or not the child had health insurance. Disability status, as one might expect, had the overall strongest effect of all health measures included in my model. As a reminder, disability status was a compilation of children with either a mental health disability, physical health disability, or special services qualification.[11] The ECLS-K did not provide a breakdown of this measure in their public dataset, prohibiting me from using separate measures for mental health and physical health disabilities. I found that the effect of disability status on reading achievement increased over time. The effect of disability status on math achievement, on the other hand, decreased after an initial moderate effect at kindergarten entry. This different pattern of effects on math and reading achievement over time was interesting, though hard to explain. The diminishing effect of disability over time in math

[11] It should be noted that my sample was limited to students who were in normal education classrooms, which meant that children with the most severe disabilities, those who had been placed in special education classrooms, were not included in my analyses.

could indicate that children with disabilities were able to adjust after kindergarten and regain ground they had lost in math achievement, while initial losses in reading achievement seemed to compound over time.

Gaps by Risk Group

The final part of my analysis of the ECLS-K data looked at the gaps between three groups of children who had different levels of regulatory behaviors. I found from my secondary analysis of ECLS-K data that children in the high risk group (those with poor regulatory behaviors) were at higher risk for a number of other common individual and contextual risk factors. The gap in reading and math achievement for children in the high-risk group compared to their average-risk peers was substantial at kindergarten entry and grew larger through each grade, ending fifth grade with a considerable gap of 0.7 sd in both reading and math. The gap between children from the high- and low-risk groups, as expected, was even wider, with a gap of 1.14 sd in reading and 1.08 sd in math.

These large gaps reinforce the challenge facing educators today. Unfortunately, I know of nothing in the school reform literature that comes close to having the effect size necessary to raise the achievement of high-risk children to even the level of the average child. Borman and colleagues (2002), for example, found that existing Comprehensive School Reform programs produced an average effect on student achievement of about 0.10, though several programs did show somewhat larger effect sizes. However, even doubling the effectiveness of these programs would still not come close to reducing the achievement gaps found in my analyses.

If our goal is to raise the achievement of high-risk children to the levels of those in even the average-risk group, whole-school and academic reforms will probably not be enough. As I found in my analyses that almost half of the full achievement gap was due to measures of children's regulatory behaviors and physical health, it seems likely that improving these areas, particularly improving children's regulatory behaviors, could lead to a further reduction in the achievement gap. A growing number of both social-emotional learning (SEL) interventions and health interventions exist in schools today. I review some of the evidence from these types of programs below.

Social-Emotional Learning (SEL) Interventions. While there are now hundreds of SEL programs throughout America, only a handful have been rigorously studied. Of the programs that have been evaluated through randomized studies, almost all of them have been shown to significantly improve social-emotional outcomes. Students in PATHS schools, a program with a weekly lesson on subjects such as emotional knowledge, emotional processes, and conflict resolution demonstrated large effects particularly on emotional knowledge. Students in PATHS schools also displayed higher interpersonal skills, social competence, self-control, lower externalizing and interpersonal skills, and more response inhibition (CPPRG, 1999; Kam, Greenberg & Walls, 2003). The Chicago School Readiness Project (CSRP) intervened in head start classrooms by training teachers on behavior management strategies, providing stress-reduction workshops and coaches to teachers, and offered mental health services to children with the highest emotional and behavioral problems. The CSRP improved externalizing problems by effect sizes of .5-.6 and improved internalizing problems by effect sizes of .6-.9 (Raver et al, 2009).

CASEL (Collaborative for Academic, Social, and Emotional Learning) recently performed three comprehensive meta-analysis reviews of the SEL literature, finding that these interventions did improve social-emotional outcomes (Payton et al., 2008). The first meta-analysis was of all types of in-school SEL interventions, the second was a review for in-school interventions that only worked with children with problems, and the third was a meta-analysis of after-school SEL interventions, which were mostly used by children with demonstrated problems. In the first overall review, the effect sizes at post-intervention on overall social emotional skills was 0.60, with effect sizes of 0.23 and 0.24 on conduct problems and positive social behavior, respectively. Effect sizes at follow up were reduced, but maintained statistical significance at 0.36, 0.15, and 0.17 respectively. CASEL also looked at level of implementation, and found that those who reported no implementation problems had, as can be expected, larger effect sizes of 0.96, 0.28 and 0.31 for socio-emotional skills, conduct problems, and social behaviors respectively. The second review of programs for children with problems found larger effect sizes than the first review with effect sizes at post of 0.77, 0.47, and 0.50 on the three outcomes, with reduced but still significant effect sizes at follow-up. The after-school review found somewhat smaller effects with effect sizes of 0.17 on conduct problems and 0.22 on social behaviors.

While all of this evidence points to large improvements in targeted social-emotional skills, there was significantly less evidence that this improvement lead directly to improved achievement. The universal reviews by CASEL did find effect sizes of 0.28 at post-test, which were maintained at follow-up for their universal review. Effects were larger for interventions that targeted only children with problems – 0.43 and 0.67 at post-test and follow up, respectively. The studies included in these meta-analyses, however, were not all as statistically rigorous as one might hope, leading me to look more closely at the evidence from randomized trials of SEL programs. Evidence from these studies, unfortunately, has been less positive, finding only limited to no success on improving academic achievement (Bierman et al, 2008; Domitrovich et al, 2007; Jones et al., in review; Raver et al, 2009). One of the reasons these interventions have found such limited success at improving achievement could be that they have focused on improving children's social skills and behavior problems, with little to no emphasis on regulation skills. One intervention that focused on developing self-regulation and executive functioning skills in children, the Tools of the Mind preschool program, did have some effect on academic achievement. Randomized trials of the Tools program have found improvements in self-regulation and executive function tasks, as well as improvements in vocabulary and early literacy skills (Barnett et al., 2008; Diamond et al., 2007).

School Health Programs. A number of in-school health-service programs attempt to help children with disabilities, poor health, problems with obesity or nutrition, no insurance, or limited medical care. A nation-wide survey found that most schools in America actually provide some form of health services, such as a counselor, nurse, psychologist or social worker, though many schools reported inadequate resources (Foster, Rollefson, Doksum, Noonan, Robinson et al., 2005). About half of the schools in the sample above also worked with community-based individuals or organizations to provide services to students. Many high-risk schools have specific programs, such as School-Based Health Centers (SBHC) that provide physical and mental health assessments, screenings, immunizations, treatment for chronic illnesses, and counseling

(Geierstander, Amaral, Mansour, & Walters, 2004; Symons, Cinelli, James & Groff, 1997).

Studies of SBHCs have found that students in these schools are more likely to use health services than students in schools without SBHCs (Armbruster, Gerstein, Fallon, 1997; Kaplan, Calonge, Guernsey, Hanraham, 1998). Few studies have looked at educational outcomes, and those that have, found mixed results. The most commonly studied outcome is absenteeism, which tends to be reduced in schools with SBHCs (e.g. Gall et al., 2000 vs. Kisker & Brown, 1996). Results are inconclusive, however, on all other academic measures (see Geierstander et al., 2004 for a review).

A number of more targeted interventions are also being used in schools, such as those targeting health outcomes such as weight management (e.g. Neumark-Sztainer, 1997; Neumark-Sztainer, Martin & Story, 2000), exercise (e.g. Jamner, Spruijt-Metz, Bassin & Cooper, 2004), home visit/family health (e.g. Mitchel-Herzfeld, Izzo, Greene, Lee, & Lowenfels, 2005; Olds, Robins, O'Brien, Luckey, Pettitt et al., 2002) and specific chronic problems, such as asthma (Lwebuga-Mukasa & Dunn-Georgiou, 2002; Taras, Write, Brennan, Campana, & Lofgren, 2004). The effectiveness of these interventions on improving targeted health outcomes are mixed, and as with the studies of SBHCs, only a handful looked at academic achievement or other academic outcomes, with little evidence of improvement in achievement for children in schools with the health interventions.

I found a number of measures of health services in the ECLS-K that were gathered from schools during the first year of the study: whether or not the school had a nurse, a psychologist, or a school counselor; whether the school offered health or social services collaboratively with other agencies; and whether or not the school offered hearing and vision screening. I included each of these in my reading and math achievement models, and found that that none of them reached levels of significance, similar to the study findings mentioned above.

Looking across all the evidence of health and social-emotional learning interventions, while it is heartening to see that these interventions, for the most part, are successful in improving the areas they are targeting, it is disappointing that these results have not translated into improved achievement. One reason for so few effects on achievement could be a lack of integration of these programs with the academic curriculum. Another reason could be that most of the areas targeted by these interventions (social skills, conflict resolution, behavior problems, and physical health) are only weakly related to academic achievement. Regulatory behaviors, the construct I found which had the strongest association with academic achievement has not generally been a focus of SEL interventions. More research is needed to determine if integrated in-school interventions training self-regulation skills would lead to improvements in children's academic achievement.

Directions for Future Research

This dissertation provided evidence that children's good health and regulatory behaviors lead to improved reading and math achievement. The strongest effect on achievement was children's regulatory behaviors, a latent factor that appears to be fairly stable over time. Further research, however, should be done to verify these results. First, the stability of children's regulatory behaviors over time should be examined more fully, using more robust statistical models as well as more precise and robust measures of regulation to determine if my findings can be confirmed. Secondly, similar analyses looking jointly at the time-varying and stable effects of the spectrum of health and social-emotional and regulatory skills measures should be performed in other longitudinal datasets, once again, in order to see if my findings from the ECLS-K will be comparable to those found in other datasets.

Additionally, more research needs to be done to better understand the pathways through which health and social-emotional skills, particularly self-regulation, affects achievement. This, of course, will require longitudinal datasets with more specific and precise measures of children's social-emotional and regulation skills over time than those available in the ECLS-K. A more refined understanding of the relationship of these skills with achievement will assist educators in identifying points of leverage for developing interventions that could lead to larger improvement on achievement than those currently available.

In the meantime, one other avenue of research that could be of more immediate use to educators is to look at the possible moderating effect of social-emotional and regulation skills on aspects of the classroom context. A moderating effect would occur if features and processes within the classroom context have a differential effect on reading and math achievement for children with different levels of social-emotional and regulations skills. Understanding these moderating effects can help educators identify features and practices in the classroom that are either protective or vulnerable for at-risk children. A few studies have examined how children's social-emotional and academic skills moderate the effect of teacher and classroom measures on reading and math achievement, finding that the warmth and closeness of the teacher-child relationship, the emotional support provided by the teacher, and the time spent in different instructional settings (child-managed vs. teacher-directed whole group) all had differential effects on achievement for children with different levels of social and academic skills (Baker, 2006; Baker, Grant & Morlock, 2008; Connor, Jakobsons, Crowe & Meadows, 2009; Hamre & Pianta, 2001; Hamre & Pianta, 2005; Morrison & Connor, 2002). More research, however, is needed to investigate the moderating effect of a range of child skills, particularly regulation skills, on a wider set of classroom and teacher measures.

Appendix

Fully specified HCM model for reading and math achievement

Level-1 Model: Time Points

$Y = \pi_0 + \pi_1$*(K Growth) + π_2*(K-G1 Growth) + π_3*(G1-G3 Growth) + π_4*(G3-G5 Growth)
+ π_6*(Regulatory Behaviors) + π_6*(Internalizing Problems) + π_7*(Health Scale)
+ π_8*(Disability Status) + π_9*(No Doctor/Dentist Visit) + π_{10}*(No Health Insurance)
+ π_{11}*(Class Size) + π_{12}*(% Minority in Class) + π_{13}*(% Disability in Class)
+ π_{14}*(% Read Below Grade Level) + π_{15}*(Class Behavior) + π_{16}*(Amount Read)
+ π_{17}*(Call home for Good Behavior) + π_{18}*(Call home for Problems) + π_{19}*(Parent Volunteers)
+ π_{20}*(# Volunteer Hours) + π_{21}*(Teacher is Minority Race)
+ π_{22}*(Teacher has Masters or higher) + π_{23}*(# Years Teaching)
+ π_{24}*(Regulatory Behaviors * K Growth) + π_{27}*(Regulatory Behaviors * K-G1 Growth)
+ π_{30}*(Regulatory Behaviors * G1-G3 Growth) + π_{33}*(Regulatory Behaviors * G3-G5 Growth)
+ π_{25}*(Health * K Growth) + π_{28}*(Health * K-G1 Growth) + π_{31}*(Health * G1-G3 Growth)
+ π_{34}*(Health * G3-G5 Growth) + π_{26}*(Disability * K Growth) + π_{29}*(Disability * K-G1 Growth)
+ π_{32}*(Disability * G1-G3 Growth) + π_{35}*(Disability * G3-G5 Growth) + e

Level-2 Model: Child Level and School Level

$\pi_0 = \theta_0 + b_{00}$
+ (β_{02})*Birth Weight + (β_{03})*Premature + (β_{04})*BMI + (β_{05})*Male + (β_{06})*Black + (β_{07})*Hispanic
+ (β_{08})*Asian + (β_{09})*Other + (β_{010})*Child Age + (β_{011})*No English in Home + (β_{012})*Repeat K
+ (β_{013})*Parents Chose School + (β_{014})*Parent believes Behavior Skills are Important for K
+ (β_{015})*Food Insecurity + (β_{016})*Home Environmental Risk Index
+ (β_{017})*Full Day Kindergarten + (γ_{01})*Northeast + (γ_{02})*South + (γ_{03})*West + (γ_{04})*Private
+ (γ_{05})*School Enrollment + (γ_{06})*% Students at Grade Level in Math/Reading
+ (γ_{07})*# Full Time Special Education teachers + (γ_{08})*# Full Time Nurses
+ (γ_{09})*Policy for Quiet/Orderly Environment in School + (γ_{010})*School Neighborhood Safety

$\pi_1 = \theta_1 + c_{10}$
+ (β_{12})*Birth Weight + (β_{13})*Premature + (β_{14})*BMI + (β_{15})*Male + (β_{16})*Black + (β_{17})*Hispanic
+ (β_{18})*Asian + (β_{19})*Other + (β_{110})*Child Age + (β_{111})*No English in Home + (β_{112})*Repeat K
+ (β_{113})*Parents Chose School + (β_{114})*Parent believes Behavior Skills are Important for K
+ (β_{115})*Food Insecurity + (β_{116})*Home Environmental Risk Index
+ (β_{117})*Full Day Kindergarten + (γ_{11})*Northeast + (γ_{12})*South + (γ_{13})*West + (γ_{14})*Private
+ (γ_{15})*School Enrollment + (γ_{16})*% Students at Grade Level in Math/Reading
+ (γ_{17})*# Full Time Special Education teachers + (γ_{18})*# Full Time Nurses
+ (γ_{19})*Policy for Quiet/Orderly Environment in School + (γ_{110})*School Neighborhood Safety

$\pi_2 = \theta_2 + b_{20}$

+ (β_{02})*Birth Weight + (β_{23})*Premature + (β_{24})*BMI + (β_{25})*Male + (β_{26})*Black + (β_{27})*Hispanic
+ (β_{28})*Asian + (β_{29})*Other + (β_{210})*Child Age + (β_{211})*No English in Home + (β_{212})*Repeat K
+ (β_{213})*Parents Chose School + (β_{214})*Parent believes Behavior Skills are Important for K
+ (β_{215})*Food Insecurity + (β_{216})*Home Environmental Risk Index
+ (β_{217})*Full Day Kindergarten + (γ_{21})*Northeast + (γ_{22})*South + (γ_{23})*West + (γ_{24})*Private
+ (γ_{25})*School Enrollment + (γ_{26})*% Students at Grade Level in Math/Reading
+ (γ_{27})*# Full Time Special Education teachers + (γ_{28})*# Full Time Nurses
+ (γ_{29})*Policy for Quiet/Orderly Environment in School + (γ_{210})*School Neighborhood Safety

$\pi_3 = \theta_3 + b_{30}$
+ (β_{32})*Birth Weight + (β_{33})*Premature + (β_{34})*BMI + (β_{35})*Male + (β_{36})*Black + (β_{37})*Hispanic
+ (β_{38})*Asian + (β_{39})*Other + (β_{310})*Child Age + (β_{311})*No English in Home + (β_{312})*Repeat K
+ (β_{313})*Parents Chose School + (β_{314})*Parent believes Behavior Skills are Important for K
+ (β_{315})*Food Insecurity + (β_{316})*Home Environmental Risk Index
+ (β_{317})*Full Day Kindergarten + (γ_{31})*Northeast + (γ_{32})*South + (γ_{33})*West + (γ_{34})*Private
+ (γ_{35})*School Enrollment + (γ_{36})*% Students at Grade Level in Math/Reading
+ (γ_{37})*# Full Time Special Education teachers + (γ_{38})*# Full Time Nurses
+ (γ_{39})*Policy for Quiet/Orderly Environment in School + (γ_{310})*School Neighborhood Safety

$\pi_4 = \theta_4 + b_{40}$
+ (β_{42})*Birth Weight + (β_{43})*Premature + (β_{44})*BMI + (β_{45})*Male + (β_{46})*Black + (β_{47})*Hispanic
+ (β_{48})*Asian + (β_{49})*Other + (β_{410})*Child Age + (β_{411})*No English in Home + (β_{412})*Repeat K
+ (β_{413})*Parents Chose School + (β_{414})*Parent believes Behavior Skills are Important for K
+ (β_{415})*Food Insecurity + (β_{416})*Home Environmental Risk Index
+ (β_{417})*Full Day Kindergarten + (γ_{41})*Northeast + (γ_{42})*South + (γ_{43})*West + (γ_{44})*Private
+ (γ_{45})*School Enrollment + (γ_{46})*% Students at Grade Level in Math/Reading
+ (γ_{47})*# Full Time Special Education teachers + (γ_{48})*# Full Time Nurses
+ (γ_{49})*Policy for Quiet/Orderly Environment in School + (γ_{410})*School Neighborhood Safety

$\pi_5 = \theta_5$
$\pi_6 = \theta_6$
$\pi_7 = \theta_7$
$\pi_8 = \theta_8$
$\pi_9 = \theta_9$
$\pi_{10} = \theta_{10}$
$\pi_{11} = \theta_{11}$
$\pi_{12} = \theta_{12}$
$\pi_{13} = \theta_{13}$
$\pi_{14} = \theta_{14}$
$\pi_{15} = \theta_{15}$
$\pi_{16} = \theta_{16}$
$\pi_{17} = \theta_{17}$
$\pi_{18} = \theta_{18}$
$\pi_{19} = \theta_{19}$
$\pi_{20} = \theta_{20}$
$\pi_{21} = \theta_{21}$
$\pi_{22} = \theta_{22}$
$\pi_{23} = \theta_{23}$
$\pi_{24} = \theta_{24}$
$\pi_{25} = \theta_{25}$
$\pi_{26} = \theta_{26}$

$\pi_{27} = \theta_{27}$
$\pi_{28} = \theta_{28}$
$\pi_{29} = \theta_{29}$
$\pi_{30} = \theta_{30}$
$\pi_{31} = \theta_{31}$
$\pi_{32} = \theta_{32}$
$\pi_{33} = \theta_{33}$
$\pi_{34} = \theta_{34}$
$\pi_{35} = \theta_{35}$

Table A.1
Unstandardized mean differences and standard errors for movers flagged to be followed versus movers not flagged at each data collection round

	Kindergarten to First Grade Movers			First Grade to Third Grade Movers			Third Grade to Fifth Grade Movers		
	mean	*(se)*		*mean*	*(se)*		*mean*	*(se)*	
Male	0.0133	(0.0136)		-0.0223	(0.0109)	*	-0.0203	(0.0098)	*
Black	0.0270	(0.0292)		-0.0260	0.0072	***	-0.0122	(0.0059)	***
Hispanic	-0.8522	(0.0401)	*	0.0488	(0.0074)	***	0.0257	(0.0065)	***
Asian	-0.0353	(0.0208)		0.0086	(0.0051)		0.0091	(0.0044)	*
Other	0.0069	(0.0181)		-0.0010	(0.0052)		-0.0037	(0.0340)	***
Socio-Economic Status	0.1771	(0.0641)	**	0.0660	(0.0147)	***	0.0102	(0.0133)	
Non-English Spoken in Home	-0.1102	(0.0335)	***	0.0341	(0.0080)	***	0.0344	(0.0062)	***
Child Repeated Kindergarten	-0.0036	(0.0255)		-0.0100	(0.0050)	*	-0.0104	(0.0043)	*
Single Parent	-0.0165	(0.0358)		-0.0491	(0.0095)	***	-0.0465	(0.0082)	***
Parent Education Expectations	-0.0463	(0.0844)		0.0876	(0.0258)	***	0.0564	(0.0232)	*
Unsafe Neighborhood	-0.0476	(0.0341)		-0.0167	(0.0096)		-0.0157	(0.0079)	*
Full Day Kindergarten	0.0189	(0.0407)		-0.0129	(0.0055)	*	0.0031	(0.0047)	
Kindergarten Class Size	-0.1398	(0.4337)		-0.0006	(0.0705)		0.1580	(0.0635)	*
Kindergarten Class Behavior	0.03489	(0.0142)	*	0.0457	(0.0149)	**	0.0349	(0.0142)	*

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

Table A.2
Unstandardized mean differences and standard errors for responders versus non-responders at each data collection round

	Kindergarten			First Grade			Third Grade			Fifth Grade		
	mean	*(se)*		*mean*	*(se)*		*mean*	*(se)*		*mean*	*(se)*	
K. Reading Achievement	-0.0540	(0.3871		-0.0539	(0.3872)		0.4758	(0.3461)		1.1485	(0.4114)	**
K. Math Achievement	0.2185	(0.3149		0.2185	(0.3149)		0.4276	(0.2946)		1.1557	(0.3712)	**
Approaches to Learning (0-4)	0.0996	(0.0260)	***	0.0996	(0.0260)	***	0.1223	(0.0231)	***	0.1158	(0.0321)	***
Self-Control (0-4)	0.0856	(0.0234)	***	0.0856	(0.0234)	***	0.1014	(0.0209)	***	0.1003	(0.0283)	***
Interpersonal Skills (0-4)	0.0685	(0.0234)	**	0.0685	(0.0234)	**	-0.0328	(0.0181)		0.0903	(0.0280)	**
Externalizing Problems (0-4)	-0.0990	(0.0229)	***	-0.0990	(0.0255)	***	-0.1148	(0.0218)	***	-0.0877	(0.0312)	**
Internalizing Problems (0-4)	-0.0453	(0.0191)	*	-0.0453	(0.0190)	*	-0.0328	(0.0181)		-0.0566	(0.0228)	*
Health Status (1-5)	0.0337	0.0157	*	0.0553	0.0355		0.0252	(0.0289)		0.0328	(0.0362)	
Child has Disability	-0.0159	0.0065	*	-0.0297	0.0148	*	-0.0449	(0.0143)	**	-0.0593	(0.0148)	***
Male	-0.0008	0.0094		-0.0220	0.0182		-0.0513	(0.0177)	**	-0.0654	(0.0229)	**
Black	0.0009	0.0079		0.0014	0.0114		0.0201	(0.0091)	*	0.0018	(0.0107)	
Hispanic	-0.0110	0.0079		0.0210	0.0123		-0.0216	(0.0115)		-0.0262	(0.0167)	
Asian	0.0072	0.0060		0.0103	0.0069		0.0150	(0.0067)	*	0.0077	(0.0106)	
Other Race	-0.0140	0.0050	**	-0.0082	0.0074		0.0102	(0.0070)		0.0007	(0.0110)	

	Kindergarten			First Grade			Third Grade			Fifth Grade		
	mean	*(se)*		*mean*	*(se)*		*mean*	*(se)*		*mean*	*(se)*	
Socio-Economic Status	0.1019	0.0244	***	-0.0332	0.0247		-0.0192	(0.0232)		-0.0172	(0.0296)	
Non-English Spoken in Home	0.0087	0.0108		0.0104	0.0113		-0.0009	(0.0103)		-0.0292	(0.0162)	
Child Repeated Kindergarten	-0.0077	0.0098		-0.0075	0.0082		-0.0160	(0.0071)	*	-0.0328	(0.0089)	***
Single Parent	-0.0407	0.0080	***	-0.0135	0.0167		0.0069	(0.0143)		-0.0404	(0.0534)	
Parent Education Expectations	0.0588	0.0266	*	0.0499	0.0419		0.0206	(0.0381)		-0.0007	(0.0176)	
Unsafe Neighborhood	0.0096	0.0098		0.0203	0.0153		-0.0009	(0.0146)		-0.0258	(0.0186)	
Full Day K.	-0.0093	0.0111		-0.0075	0.0100		0.0005	(0.0089)		-0.0004	(0.0109)	
K. Class Size	0.0789	0.1394		0.1666	0.1447		0.1135	(0.1355)		0.2089	(0.1912)	
K. Class Behavior	0.0289	0.0209		0.0281	0.0284		0.0483	(0.0239)	*	-0.0048	(0.0310)	

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

Table A.3
Randomization inference omnibus test testing for balance on child and teacher measures across five propensity strata

	Average Regulatory Behaviors=0	Average Regulatory Behaviors=1	Standardized Difference	Z
Male	0.5797	0.58499	0.0108	1.1242
Black	0.1458	0.1428	-0.0096	-0.8356
Hispanic	0.1971	0.1992	0.0054	0.4793
Asian	0.0574	0.0590	0.0064	0.6386
Other	0.0558	0.0567	0.0039	0.3382
Age	78.2771	78.2667	-0.0023	-0.2098
No English in Home	0.1586	0.1607	0.0056	0.5059
Repeat Kindergarten	0.0505	0.0483	-0.0108	-0.9246
Parent Believes Behavior Skills are Important for School	-0.0497	-0.0494	0.0003	0.0312
Parent Chose School	0.3282	0.3278	-0.0009	-0.0859
SES	-0.0902	-0.0798	0.0105	1.0241
Single Parent	0.2367	0.2341	-0.0068	-0.6184
Parent Expectations	3.0650	3.0723	0.0067	0.5984
Does Educational Activities in Home	-0.0537	-0.0518	0.0020	0.1779
Amount Reads Picture Books in Home	0.8563	0.8562	-0.0005	-0.0401
Amount Reads in Home	0.6668	0.6640	-0.0061	-0.5507
Bad Parent Mental Health	0.2878	0.2868	-0.0022	-0.2062
Good Parent Mental Health	0.3682	0.3658	-0.0051	-0.4632
Argue in Home	-0.0131	-0.0114	0.0018	0.1598
Poor Home Environment	0.2528	0.2522	-0.0013	-0.1180
Good Home Environment	0.3032	0.3028	-0.0007	-0.0622
Bedtime varies	0.1205	0.1201	-0.0015	-0.1341
Neighborhood Safety	0.2708	0.2702	-0.0015	-0.1380
Full day Kindergarten	0.5669	0.5670	0.0002	0.0162
Class Size	23.125	23.1275	0.0003	0.0280

	Average Regulatory Behaviors=0	Average Regulatory Behaviors=1	Standardized Difference	Z
1-20% Minorities in class	0.4118	0.4135	0.0036	0.3249
21-100% Minorities in Class	0.2504	0.2498	-0.0016	-0.1436
% Read Below Grade Level	0.1652	0.1645	-0.0056	-0.4950
# Hours Volunteers in class	3.2660	3.2886	0.0042	0.3930
Rating of Class Behavior	0.4940	0.5012	0.0087	0.8166
Teacher calls home for good behavior	0.7887	0.7906	0.0046	0.4176
Teacher calls home for problems	0.7480	0.7458	-0.0050	-0.4719
Parent Volunteered in Classroom	0.4864	0.4861	-0.0006	-0.0556
Amount of Time in Whole Class Instruction	0.5410	0.5430	0.0022	0.2022
Amount of Time in Small Group Instruction	-0.0330	-0.0308	0.0026	0.2335
Amount of time in Individual Activities	-0.6412	-0.6392	0.0028	0.2504
Amount of time in Child-Directed Activities	-0.3558	-0.3541	0.0023	0.2098
1-100 minutes of reading achievement groups a week	0.4368	0.4339	-0.0059	-0.5344
100+ minutes of reading achievement groups a week	0.1944	0.1969	0.0064	0.5730
Overall Test of Balance				
χ^2	8.59			
df	39			
p-value	1			

Table A.4
Full list of coefficients and standard errors from HCM model of reading achievement using regulatory behaviors either as a stable or time-varying covariate

	Reading IRT Scale Score *with stable regulatory behaviors*		**Reading IRT Scale Score** *with time-varying regulatory behaviors*	
	β	*SE*	*β*	*SE*
π_0: Achievement at Fall Kindergarten				
θ_0: Initial Fall K Achievement	28.422	0.573 **	28.522	0.573**
β_{01}: Average Regulatory Behaviors	1.830	0.114 **		
β_{02}: Birth Weight	0.006	0.003 ^	0.007	0.003*
β_{03}: Premature	-0.049	0.321	-0.055	0.324
β_{04}: Average BMI	-0.279	0.102 **	-0.310	0.102**
β_{05}: Food Insecurity	-0.935	0.402 *	-1.001	0.404*
β_{06}: Male	-0.298	0.218	-0.642	0.212**
β_{07}: Black	0.192	0.385	-0.107	0.386
β_{08}: Hispanic	-1.552	0.358 **	-1.540	0.360**
β_{09}: Asian	3.733	0.605 **	4.064	0.599**
β_{010}: Other	-0.150	0.475	-0.170	0.478
β_{011}: Age	0.327	0.024 **	0.327	0.024**
β_{012}: No English in Home	-2.117	0.408 **	-1.983	0.412**
β_{013}: Repeat Kindergarten	1.510	0.550 **	1.410	0.557*
β_{014}: Parent Chooses School	0.710	0.257 **	0.716	0.258**
β_{015}: Behavior skills important for K	0.500	0.100 **	0.481	0.101**
β_{016}: Average Home Risk Index	-0.904	0.116 **	-0.965	0.116**
β_{017}: Full Day Kindergarten	0.501	0.233 *	0.483	0.234*
γ_{01}: Northeast	0.993	0.317 **	1.065	0.318**
γ_{02}: South	0.510	0.297 ^	0.589	0.299*
γ_{03}: West	1.528	0.317 **	1.579	0.318**
γ_{04}: Private School	1.654	0.368 **	1.720	0.369**
γ_{05}: School Enrollment	0.172	0.099 ^	0.188	0.099^
γ_{06}: %Students below grade in Math	0.023	0.005 **	0.023	0.005**
γ_{07}: Special Ed FTE	0.045	0.040	0.037	0.040
γ_{08}: Nurse FTE	0.016	0.118	0.019	0.119
γ_{09}: Policy for Quiet/Orderly Environment	0.191	0.196	0.204	0.196
γ_{010}: School Neighborhood Safety	0.068	0.120	0.085	0.121
π_1: Linear Growth from Fall Kindergarten to Spring Kindergarten				
θ_1: Fall K to Spring K Increase in Achievement	10.464	0.431 **	10.587	0.432**
β_{11}: Average Regulatory Behaviors	1.180	0.088 **		
β_{12}: Birth Weight	0.003	0.003	0.003	0.003
β_{13}: Premature	0.008	0.258	0.001	0.258
β_{14}: Average BMI	-0.053	0.081	-0.074	0.081

	Reading IRT Scale Score *with stable regulatory behaviors*		Reading IRT Scale Score *with time-varying regulatory behaviors*	
	β	*SE*	*β*	*SE*
β_{15}: Food Insecurity	-0.084	0.354	-0.120	0.354
β_{16}: Male	-0.037	0.170	-0.276	0.166^
β_{17}: Black	-0.775	0.317 *	-1.015	0.316**
β_{18}: Hispanic	-0.472	0.282 ^	-0.463	0.283
β_{19}: Asian	0.898	0.427 *	1.094	0.426*
β_{110}: Other	0.344	0.402	0.304	0.404
β_{111}: Age	0.018	0.019	0.020	0.019
β_{112}: No English in Home	0.061	0.361	0.145	0.362
β_{113}: Repeat Kindergarten	-1.609	0.544 **	-1.689	0.544**
β_{114}: Parent Chooses School	0.603	0.226 **	0.605	0.225**
β_{115}: Behavior skills important for K	0.265	0.081 **	0.252	0.081**
β_{116}: Average Home Risk Index	-0.092	0.089	-0.135	0.089
β_{117}: Full Day Kindergarten	1.536	0.322 **	1.533	0.323**
γ_{11}: Northeast	-0.343	0.406	-0.286	0.409
γ_{12}: South	0.579	0.347 ^	0.643	0.348^
γ_{13}: West	1.273	0.402 **	1.316	0.403**
γ_{14}: Private School	0.033	0.408	0.105	0.409
γ_{15}: School Enrollment	-0.067	0.128	-0.048	0.128
γ_{16}: %Students below grade in Math	0.001	0.006	0.001	0.006
γ_{17}: Special Ed FTE	0.038	0.048	0.029	0.048
γ_{18}: Nurse FTE	0.005	0.151	0.009	0.151
γ_{19}: Policy for Quiet/Orderly Environment	0.479	0.250 ^	0.483	0.253^
γ_{110}: School Neighborhood Safety	-0.061	0.153	-0.042	0.154
π_2: Linear Growth from Spring Kindergarten to Spring Grade 1				
θ_2: Spring K to Spring G1 Increase in Achievement	31.532	0.516 **	31.806	0.518**
β_{21}: Average Regulatory Behaviors	2.759	0.162 **		
β_{22}: Birth Weight	0.002	0.005	0.003	0.005
β_{23}: Premature	0.027	0.408	0.036	0.411
β_{24}: Average BMI	0.050	0.147	0.014	0.148
β_{25}: Food Insecurity	-0.909	0.606	-1.025	0.602^
β_{26}: Male	0.142	0.309	-0.412	0.307
β_{27}: Black	-1.639	0.552 **	-2.107	0.548**
β_{28}: Hispanic	-1.210	0.551 *	-1.232	0.552*
β_{29}: Asian	0.775	0.796	1.059	0.799
β_{210}: Other	-0.693	0.722	-0.749	0.727
β_{211}: Age	0.062	0.035 ^	0.069	0.035*
β_{212}: No English in Home	-1.146	0.566 *	-1.046	0.567^
β_{213}: Repeat Kindergarten	-3.652	1.287 **	-3.892	1.287**

β_{214}: Parent Chooses School	0.222	0.426	0.245	0.431
	Reading IRT Scale Score *with stable regulatory behaviors*		**Reading IRT Scale Score** *with time-varying regulatory behaviors*	
	β	*SE*	*β*	*SE*
β_{215}: Behavior skills important for K	0.241	0.154	0.239	0.154
β_{216}: Average Home Risk Index	-0.598	0.165 **	-0.701	0.165**
β_{217}: Full Day Kindergarten	-0.849	0.344 *	-0.844	0.345*
γ_{21}: Northeast	0.605	0.489	0.804	0.494
γ_{22}: South	1.285	0.444 **	1.395	0.447**
γ_{23}: West	0.455	0.488	0.498	0.493
γ_{24}: Private School	0.557	0.633	0.559	0.633
γ_{25}: School Enrollment	0.120	0.160	0.158	0.161
γ_{26}: %Students below grade in Math	0.015	0.007 *	0.016	0.007*
γ_{27}: Special Ed FTE	-0.079	0.057	-0.084	0.057
γ_{28}: Nurse FTE	0.088	0.195	0.113	0.196
γ_{29}: Policy for Quiet/Orderly Environment	-0.146	0.305	-0.157	0.306
γ_{210}: School Neighborhood Safety	0.307	0.187	0.283	0.189
π_3: Linear Growth from Spring Grade 1 to Spring Grade 3				
θ_3: Spring G1 to Spring G3 Increase in Achievement	47.564	0.669 **	47.600	0.665**
β_{31}: Average Regulatory Behaviors	1.107	0.197 **		
β_{32}: Birth Weight	-0.008	0.006	-0.009	0.006
β_{33}: Premature	0.715	0.556	0.640	0.557
β_{34}: Average BMI	0.321	0.180 ^	0.343	0.181^
β_{35}: Food Insecurity	-0.869	0.712	-0.782	0.718
β_{36}: Male	0.425	0.365	0.412	0.363
β_{37}: Black	-4.935	0.689 **	-4.939	0.692**
β_{38}: Hispanic	-2.128	0.675 **	-2.196	0.679**
β_{39}: Asian	-5.182	0.891 **	-5.235	0.890**
β_{310}: Other	-3.801	0.985 **	-3.846	0.987**
β_{311}: Age	-0.171	0.041 **	-0.170	0.042**
β_{312}: No English in Home	-1.877	0.645 **	-1.916	0.645**
β_{313}: Repeat Kindergarten	0.294	1.480	0.190	1.486
β_{314}: Parent Chooses School	-0.078	0.483	-0.112	0.487
β_{315}: Behavior skills important for K	-0.055	0.176	-0.035	0.176
β_{316}: Average Home Risk Index	-1.228	0.196 **	-1.243	0.197**
β_{317}: Full Day Kindergarten	-0.588	0.415	-0.642	0.415
γ_{31}: Northeast	-1.237	0.558 *	-1.333	0.559*
γ_{32}: South	-1.198	0.537 *	-1.271	0.539*
γ_{33}: West	-1.453	0.633 *	-1.548	0.633*
γ_{34}: Private School	0.392	0.805	0.436	0.811
γ_{35}: School Enrollment	0.015	0.199	0.020	0.199

γ_{36}: %Students below grade in Math	0.016	0.009 ^	0.016	0.009^
γ_{37}: Special Ed FTE	0.122	0.076	0.131	0.076^
	Reading IRT Scale Score *with stable regulatory behaviors*		**Reading IRT Scale Score** *with time-varying regulatory behaviors*	
	β	*SE*	β	*SE*
γ_{38}: Nurse FTE	-0.156	0.216	-0.177	0.216
γ_{39}: Policy for Quiet/Orderly Environment	-0.580	0.371	-0.574	0.375
γ_{310}: School Neighborhood Safety	-0.073	0.232	-0.075	0.234
π_4: Linear Growth from Spring Grade 3 to Spring Grade 5				
θ_4: Spring G3 to Spring G5 Increase in Achievement	20.127	0.520 **	20.046	0.523**
β_{41}: Average Regulatory Behaviors	-0.880	0.159 **		
β_{42}: Birth Weight	0.001	0.005	0.000	0.005
β_{43}: Premature	-0.266	0.417	-0.246	0.417
β_{44}: Average BMI	0.223	0.150	0.229	0.150
β_{45}: Food Insecurity	0.894	0.748	0.940	0.752
β_{46}: Male	0.569	0.312 ^	0.851	0.311**
β_{47}: Black	1.107	0.610 ^	1.489	0.614*
β_{48}: Hispanic	1.474	0.494 **	1.509	0.496**
β_{49}: Asian	0.692	0.712	0.407	0.711
β_{410}: Other	1.569	0.795 *	1.559	0.802^
β_{411}: Age	-0.145	0.033 **	-0.137	0.034**
β_{412}: No English in Home	1.176	0.606 ^	1.075	0.608^
β_{413}: Repeat Kindergarten	-0.010	0.936	0.083	0.908
β_{414}: Parent Chooses School	0.419	0.398	0.398	0.398
β_{415}: Behavior skills important for K	-0.139	0.146	-0.136	0.146
β_{416}: Average Home Risk Index	0.201	0.162	0.260	0.162
β_{417}: Full Day Kindergarten	-0.797	0.341 *	-0.820	0.342*
γ_{41}: Northeast	1.205	0.456 **	1.154	0.457*
γ_{42}: South	0.163	0.412	0.133	0.413
γ_{43}: West	-0.398	0.470	-0.438	0.473
γ_{44}: Private School	-0.364	0.564	-0.560	0.565
γ_{45}: School Enrollment	-0.348	0.161 *	-0.401	0.161*
γ_{46}: %Students below grade in Math	-0.019	0.007 *	-0.021	0.007**
γ_{47}: Special Ed FTE	0.079	0.066	0.083	0.066
γ_{48}: Nurse FTE	-0.173	0.188	-0.170	0.188
γ_{49}: Policy for Quiet/Orderly Environment	-0.267	0.300	-0.236	0.301
γ_{410}: School Neighborhood Safety	-0.145	0.179	-0.143	0.179
π_5 – π_{35}: Time-Varying Covariates				
θ_5: Regulatory Behaviors			1.505	0.104**

θ_6: Internalizing Problem Behavior	-0.385	0.057 **	-0.337	0.058**
θ_7: Health Scale	-0.326	0.101 **	-0.354	0.101**
θ_8: Disability Status	-0.714	0.303 *	-0.693	0.304*
	Reading IRT Scale Score *with stable regulatory behaviors*		**Reading IRT Scale Score** *with time-varying regulatory behaviors*	
	β	*SE*	*β*	*SE*
θ_9: No Doctor/Dentist Visit in past year	0.335	0.158 *	0.333	0.159*
θ_{10}: No Health Insurance	-0.065	0.205	-0.066	0.206
θ_{11}: Class Size	-0.029	0.058	-0.012	0.058
θ_{12}: % Minority in Class	-0.667	0.075 **	-0.670	0.075**
θ_{13}: % Disability in class	-0.537	0.061 **	-0.546	0.061**
θ_{14}: % Read below grade level in class	-0.857	0.070 **	-0.872	0.071**
θ_{15}: Teacher rating of Class Behavior	0.122	0.058 *	0.170	0.058**
θ_{16}: Amount Time Spent in Reading Instruction	0.250	0.059 **	0.248	0.059**
θ_{17}: Teacher calls home for good behavior	0.841	0.213 **	1.054	0.213**
θ_{18}: Teacher calls home for Bad Behavior	-0.806	0.194 **	-1.125	0.196**
θ_{19}: Parent volunteered in class	0.956	0.128 **	1.083	0.128**
θ_{20}: # volunteer hours in class	0.129	0.065 *	0.141	0.065*
θ_{21}: Teacher of Minority Race	-0.136	0.167	-0.165	0.168
θ_{22}: Teacher has masters degree or higher	-0.217	0.218	-0.250	0.218
θ_{23}: # Years teaching experience	0.146	0.044 **	0.163	0.044**
θ_{24}: Regulatory Behaviors * K Growth Rate			0.961	0.086**
θ_{27}: Regulatory Behaviors * K-G1 Growth Rate			-0.134	0.096
θ_{30}: Regulatory Behaviors * Spring G3 Growth Rate			-0.377	0.084**
θ_{33}: Regulatory Behaviors * G3-G5 Growth Rate			-0.639	0.053**
θ_{25}: Health Scale * K Growth Rate	-0.114	0.088	-0.137	0.088
θ_{28}: Health Scale * K-G1 Growth Rate	-0.230	0.158	-0.238	0.158
θ_{31}: Health Scale * G1-G3 Growth Rate	0.066	0.165	0.093	0.165
θ_{34}: Health Scale * G3-G5 Growth Rate	0.278	0.086 **	0.296	0.087**
θ_{26}: Disability Status * K Growth Rate	-0.736	0.237 **	-0.747	0.239**
θ_{29}: Disability Status * K-G1 Growth Rate	-1.089	0.451 *	-1.141	0.449*
θ_{32}: Disability Status * G1-G3 Growth Rate	1.105	0.513 *	1.061	0.515*
θ_{35}: Disability Status * G3-G5 Growth Rate	0.362	0.256	0.378	0.258

Child Level Variance components						
b_{00}: Fall Kindergarten Intercept (Level 2) Random Child Effect	8.980	80.645	**	9.013	81.228	**
b_{20}: Spring K to Spring Growth Rate Random Child Effect	12.553	157.588	**	12.650	160.032	**
b_{30}: Spring G1 to Spring G3 Growth Rate Random Child Effect	15.139	229.175	**	15.198	230.966	**

	Reading IRT Scale Score *with stable regulatory behaviors*			**Reading IRT Scale Score** *with time-varying regulatory behaviors*		
	β	*SE*		β	*SE*	
b_{40}: Spring G3 to Spring G5 Growth Rate Random Child Effect	10.715	114.806	**	10.765	115.892	**
e: Level-1 error	5.593	31.286	**	5.610	31.468	**
School Level Variance Components						
c_{10}: Fall K to Spring K Growth Rate Random School Effect	2.803	7.855	**	2.836	8.04303	**

% Variance in Error and Random Effects Explained by Model		
b_{00}: Fall Kindergarten Intercept (Level 2) Random Child Effect	24.05%	23.50%
b_{20}: Spring K to Spring Growth Rate Random Child Effect	16.35%	15.05%
b_{30}: Spring G1 to Spring G3 Growth Rate Random Child Effect	4.97%	4.22%
b_{40}: Spring G3 to Spring G5 Growth Rate Random Child Effect	0.00%	0.00%
c_{10}: Fall K to Spring K Growth Rate Random School Effect	30.95%	29.30%
e: Level 1 Error	3.75%	3.19%

~ $p < .10$ * $p < .05$ ** $p < .01$

Note: Coefficients in this table are unstandardized

Table A.5
Full list of coefficients and standard errors from HCM model of math achievement using regulatory behaviors either as a stable or time-varying covariate

	Math IRT Scale Score *with stable regulatory behaviors*		**Math IRT Scale Score** *with time-varying regulatory behaviors*	
	β	*SE*	*β*	*SE*
π_0: Achievement at Fall Kindergarten				
θ_0: Initial Fall K Achievement	23.247	0.460 **	23.430	0.461**
β_{01}: Average Regulatory Behaviors	2.054	0.091 **		
β_{02}: Birth Weight	0.012	0.003 **	0.013	0.003**
β_{03}: Premature	-0.192	0.246	-0.204	0.250
β_{04}: Average BMI	-0.231	0.081 **	-0.272	0.081**
β_{05}: Food Insecurity	-0.597	0.327 ^	-0.682	0.329*
β_{06}: Male	1.385	0.168 **	0.940	0.166**
β_{07}: Black	-1.489	0.302 **	-1.869	0.303**
β_{08}: Hispanic	-2.473	0.305 **	-2.470	0.308**
β_{09}: Asian	1.842	0.389 **	2.226	0.391**
β_{010}: Other	-1.330	0.386 **	-1.385	0.389**
β_{011}: Age	0.441	0.019 **	0.445	0.019**
β_{012}: No English in Home	-1.888	0.310 **	-1.739	0.313**
β_{013}: Repeat Kindergarten	-0.418	0.412	-0.578	0.415
β_{014}: Parent Chooses School	0.564	0.213 **	0.568	0.215**
β_{015}: Behavior skills important for K	0.363	0.081 **	0.344	0.081**
β_{016}: Average Home Risk Index	-0.932	0.094 **	-1.019	0.094**
β_{017}: Full Day Kindergarten	0.317	0.188 ^	0.282	0.189
γ_{01}: Northeast	0.276	0.255	0.369	0.256
γ_{02}: South	-0.230	0.233	-0.136	0.234
γ_{03}: West	0.800	0.262 **	0.864	0.265**
γ_{04}: Private School	1.770	0.272 **	1.853	0.273**
γ_{05}: School Enrollment	0.138	0.077 ^	0.157	0.077*
γ_{06}: %Students below grade in Math	0.018	0.004 **	0.018	0.004**
γ_{07}: Special Ed FTE	0.006	0.030	-0.003	0.030
γ_{08}: Nurse FTE	0.080	0.090	0.086	0.091
γ_{09}: Policy for Quiet/Orderly Environment	-0.075	0.152	-0.060	0.151
γ_{010}: School Neighborhood Safety	0.089	0.092	0.113	0.092
π_1: Linear Growth from Fall Kindergarten to Spring Kindergarten				
θ_1: Fall K to Spring K Increase in Achievement	9.792	0.356 **	9.896	0.355**
β_{11}: Average Regulatory Behaviors	1.133	0.071 **		
β_{12}: Birth Weight	-0.002	0.002	-0.002	0.002
β_{13}: Premature	-0.125	0.205	-0.132	0.205
β_{14}: Average BMI	0.080	0.065	0.062	0.065

	Math IRT Scale Score *with stable regulatory behaviors*		Math IRT Scale Score *with time-varying regulatory behaviors*	
	β	*SE*	*β*	*SE*
β_{15}: Food Insecurity	-0.009	0.267	-0.041	0.268
β_{16}: Male	1.113	0.133 **	0.897	0.131**
β_{17}: Black	-1.474	0.268 **	-1.690	0.267**
β_{18}: Hispanic	-0.743	0.251 **	-0.732	0.251**
β_{19}: Asian	0.098	0.347	0.288	0.347
β_{110}: Other	-0.355	0.332	-0.388	0.334
β_{111}: Age	0.051	0.016 **	0.052	0.016**
β_{112}: No English in Home	-0.507	0.274 ^	-0.426	0.272
β_{113}: Repeat Kindergarten	-0.728	0.363 *	-0.794	0.360*
β_{114}: Parent Chooses School	-0.019	0.175	-0.017	0.176
β_{115}: Behavior skills important for K	0.081	0.065	0.067	0.065
β_{116}: Average Home Risk Index	-0.134	0.073 ^	-0.172	0.073*
β_{117}: Full Day Kindergarten	0.869	0.207 **	0.867	0.209**
γ_{11}: Northeast	-0.886	0.298 **	-0.837	0.301**
γ_{12}: South	-0.082	0.267	-0.023	0.269
γ_{13}: West	0.324	0.312	0.365	0.312
γ_{14}: Private School	0.296	0.304	0.361	0.306
γ_{15}: School Enrollment	0.042	0.098	0.058	0.099
γ_{16}: %Students below grade in Math	0.006	0.005	0.006	0.005
γ_{17}: Special Ed FTE	0.020	0.037	0.011	0.038
γ_{18}: Nurse FTE	-0.057	0.117	-0.054	0.118
γ_{19}: Policy for Quiet/Orderly Environment	0.248	0.197	0.254	0.198
γ_{110}: School Neighborhood Safety	-0.084	0.116	-0.066	0.116
π_2: Linear Growth from Spring Kindergarten to Spring Grade 1				
θ_2: Spring K to Spring G1 Increase in Achievement	24.490	0.398 **	24.635	0.400**
β_{21}: Average Regulatory Behaviors	1.317	0.118 **		
β_{22}: Birth Weight	0.003	0.004	0.004	0.004
β_{23}: Premature	0.058	0.309	0.063	0.310
β_{24}: Average BMI	-0.181	0.107 ^	-0.196	0.107^
β_{25}: Food Insecurity	0.175	0.460	0.113	0.463
β_{26}: Male	1.951	0.222 **	1.670	0.219**
β_{27}: Black	-2.563	0.404 **	-2.769	0.405**
β_{28}: Hispanic	-0.487	0.381	-0.501	0.383
β_{29}: Asian	-1.380	0.540 *	-1.288	0.540*
β_{210}: Other	-1.472	0.528 **	-1.500	0.531**
β_{211}: Age	-0.015	0.026	-0.010	0.026
β_{212}: No English in Home	-0.008	0.427	0.025	0.429
β_{213}: Repeat Kindergarten	-1.539	0.702 *	-1.694	0.707*

	Math IRT Scale Score *with stable regulatory behaviors*		Math IRT Scale Score *with time-varying regulatory behaviors*	
β_{214}: Parent Chooses School	-0.054	0.277	-0.035	0.278
	β	*SE*	β	*SE*
β_{215}: Behavior skills important for K	0.008	0.108	0.016	0.108
β_{216}: Average Home Risk Index	-0.427	0.121 **	-0.479	0.121**
β_{217}: Full Day Kindergarten	-0.441	0.262 ^	-0.430	0.261^
γ_{21}: Northeast	-0.636	0.344 ^	-0.518	0.346
γ_{22}: South	1.243	0.337 **	1.292	0.337**
γ_{23}: West	-0.812	0.365 *	-0.814	0.366*
γ_{24}: Private School	-0.878	0.369 *	-0.928	0.368*
γ_{25}: School Enrollment	0.293	0.114 *	0.309	0.114**
γ_{26}: %Students below grade in Math	0.003	0.005	0.003	0.005
γ_{27}: Special Ed FTE	-0.109	0.042 *	-0.107	0.042*
γ_{28}: Nurse FTE	0.188	0.130	0.205	0.130
γ_{29}: Policy for Quiet/Orderly Environment	-0.705	0.229 **	-0.720	0.230**
γ_{210}: School Neighborhood Safety	0.029	0.133	-0.002	0.133
π_3: Linear Growth from Spring Grade 1 to Spring Grade 3				
θ_3: Spring G1 to Spring G3 Increase in Achievement	32.644	0.493 **	32.780	0.497**
β_{31}: Average Regulatory Behaviors	1.783	0.149 **		
β_{32}: Birth Weight	0.005	0.005	0.005	0.005
β_{33}: Premature	0.517	0.399	0.449	0.401
β_{34}: Average BMI	0.145	0.134	0.141	0.134
β_{35}: Food Insecurity	-1.459	0.550 **	-1.429	0.549**
β_{36}: Male	2.978	0.275 **	2.719	0.273**
β_{37}: Black	-3.757	0.577 **	-3.976	0.581**
β_{38}: Hispanic	-1.325	0.469 **	-1.399	0.468**
β_{39}: Asian	1.260	0.659 ^	1.354	0.657*
β_{310}: Other	-1.259	0.631 *	-1.360	0.635*
β_{311}: Age	-0.206	0.032 **	-0.200	0.032**
β_{312}: No English in Home	-0.260	0.515	-0.249	0.517
β_{313}: Repeat Kindergarten	-0.512	0.869	-0.726	0.876
β_{314}: Parent Chooses School	0.428	0.373	0.402	0.377
β_{315}: Behavior skills important for K	0.081	0.136	0.096	0.136
β_{316}: Average Home Risk Index	-1.007	0.157 **	-1.076	0.157**
β_{317}: Full Day Kindergarten	-0.318	0.309	-0.385	0.310
γ_{31}: Northeast	0.096	0.424	0.075	0.425
γ_{32}: South	-0.098	0.394	-0.118	0.395
γ_{33}: West	0.049	0.479	-0.005	0.478
γ_{34}: Private School	-1.939	0.471 **	-1.870	0.473**
γ_{35}: School Enrollment	0.058	0.143	0.075	0.143

	Math IRT Scale Score *with stable regulatory behaviors*		Math IRT Scale Score *with time-varying regulatory behaviors*	
γ_{36}: %Students below grade in Math	0.004	0.007	0.004	0.007
γ_{37}: Special Ed FTE	0.095	0.057 ^	0.099	0.057^
	β	*SE*	*β*	*SE*
γ_{38}: Nurse FTE	-0.148	0.172	-0.159	0.173
γ_{39}: Policy for Quiet/Orderly Environment	-0.039	0.272	-0.033	0.274
γ_{310}: School Neighborhood Safety	0.247	0.180	0.253	0.182
π_4: Linear Growth from Spring Grade 3 to Spring Grade 5				
θ_4: Spring G3 to Spring G5 Increase in Achievement	20.152	0.550 **	20.076	0.550**
β_{41}: Average Regulatory Behaviors	-0.057	0.129		
β_{42}: Birth Weight	0.004	0.004	0.003	0.004
β_{43}: Premature	0.381	0.328	0.391	0.329
β_{44}: Average BMI	0.121	0.119	0.125	0.119
β_{45}: Food Insecurity	0.514	0.475	0.551	0.476
β_{46}: Male	0.426	0.249 ^	0.639	0.249*
β_{47}: Black	0.130	0.572	0.445	0.576
β_{48}: Hispanic	1.182	0.414 **	1.220	0.415**
β_{49}: Asian	1.748	0.700 *	1.540	0.697*
β_{410}: Other	1.519	0.582 **	1.521	0.583**
β_{411}: Age	-0.265	0.028 **	-0.259	0.028**
β_{412}: No English in Home	1.080	0.492 *	1.006	0.493*
β_{413}: Repeat Kindergarten	0.198	0.832	0.273	0.805
β_{414}: Parent Chooses School	0.199	0.357	0.181	0.360
β_{415}: Behavior skills important for K	0.029	0.117	0.030	0.117
β_{416}: Average Home Risk Index	0.150	0.134	0.195	0.134
β_{417}: Full Day Kindergarten	-0.151	0.323	-0.169	0.324
γ_{41}: Northeast	1.253	0.405 **	1.218	0.407**
γ_{42}: South	-0.360	0.468	-0.381	0.470
γ_{43}: West	0.141	0.509	0.115	0.510
γ_{44}: Private School	-0.087	0.541	-0.243	0.539
γ_{45}: School Enrollment	-0.518	0.131 **	-0.560	0.131**
γ_{46}: %Students below grade in Math	-0.002	0.006	-0.004	0.006
γ_{47}: Special Ed FTE	0.054	0.056	0.057	0.056
γ_{48}: Nurse FTE	-0.041	0.155	-0.041	0.156
γ_{49}: Policy for Quiet/Orderly Environment	0.107	0.276	0.137	0.277
γ_{410}: School Neighborhood Safety	-0.272	0.172	-0.271	0.174
π_5 – π_{35}: Time-Varying Covariates				
θ_5: Regulatory Behaviors			1.576	0.083**

θ_6: Internalizing Problem Behavior	-0.373	0.048 **	-0.298	0.050**
θ_7: Health Scale	-0.334	0.076 **	-0.367	0.076**
θ_8: Disability Status	-1.068	0.222 **	-1.062	0.223**
	Math IRT Scale Score *with stable regulatory behaviors*		**Math IRT Scale Score** *with time-varying regulatory behaviors*	
	β	*SE*	β	*SE*
θ_9: No Doctor/Dentist Visit in past year	0.042	0.137	0.042	0.137
θ_{10}: No Health Insurance	-0.192	0.174	-0.195	0.174
θ_{11}: Class Size	-0.056	0.047	-0.041	0.047
θ_{12}: % Minority in Class	-0.408	0.064 **	-0.416	0.065**
θ_{13}: % Disability in class	-0.342	0.049 **	-0.349	0.049**
θ_{14}: % Read below grade level in class	-0.478	0.057 **	-0.486	0.057**
θ_{15}: Teacher rating of Class Behavior	0.202	0.048 **	0.241	0.048**
θ_{16}: Amount Time Spent in Reading Instruction	0.116	0.048 *	0.115	0.048*
θ_{17}: Teacher calls home for good behavior	0.570	0.178 **	0.763	0.181**
θ_{18}: Teacher calls home for Bad Behavior	-0.648	0.168 **	-0.922	0.174**
θ_{19}: Parent volunteered in class	0.663	0.099 **	0.781	0.099**
θ_{20}: # volunteer hours in class	0.088	0.052 ^	0.096	0.052^
θ_{21}: Teacher of Minority Race	-0.170	0.138	-0.198	0.139
θ_{22}: Teacher has masters degree or higher	0.131	0.183	0.107	0.184
θ_{23}: # Years teaching experience	-0.056	0.035	-0.040	0.036
θ_{24}: Regulatory Behaviors * K Growth Rate			0.954	0.070**
θ_{27}: Regulatory Behaviors * K-G1 Growth Rate			-0.767	0.074**
θ_{30}: Regulatory Behaviors * Spring G3 Growth Rate			-0.074	0.061
θ_{33}: Regulatory Behaviors * G3-G5 Growth Rate			-0.355	0.043**
θ_{25}: Health Scale * K Growth Rate	-0.052	0.067	-0.074	0.067
θ_{28}: Health Scale * K-G1 Growth Rate	-0.095	0.108	-0.085	0.108
θ_{31}: Health Scale * G1-G3 Growth Rate	0.188	0.112 ^	0.201	0.112^
θ_{34}: Health Scale * G3-G5 Growth Rate	0.003	0.068	0.018	0.068
θ_{26}: Disability Status * K Growth Rate	-0.461	0.197 *	-0.463	0.198*
θ_{29}: Disability Status * K-G1 Growth Rate	-0.737	0.315 *	-0.773	0.316*
θ_{32}: Disability Status * G1-G3 Growth Rate	1.303	0.326 **	1.237	0.327**
θ_{35}: Disability Status * G3-G5 Growth Rate	0.320	0.205	0.345	0.206^

	Math IRT Scale Score *with stable regulatory behaviors*			Math IRT Scale Score *with time-varying regulatory behaviors*		
	β	*SE*		*β*	*SE*	
Child Level Variance components						
b_{00}: Fall Kindergarten Intercept (Level 2) Random Child Effect	7.212	52.01513	**	7.284	53.061	**
b_{20}: Spring K to Spring Growth Rate Random Child Effect	8.773	76.95896	**	8.798	77.410	**
b_{30}: Spring G1 to Spring G3 Growth Rate Random Child Effect	11.463	131.40818	**	11.504	132.347	**
b_{40}: Spring G3 to Spring G5 Growth Rate Random Child Effect	9.000	81.00574	**	9.042	81.749	**
e: Level-1 error	4.537	20.58589	**	4.553	20.726	**
School Level Variance Components						
c_{10}: Fall K to Spring K Growth Rate Random School Effect	2.127	4.523	**	2.148	4.614	**

% Variance in Error and Random Effects Explained by Model		
b_{00}: Fall Kindergarten Intercept (Level 2) Random Child Effect	35.49%	13219%
b_{20}: Spring K to Spring Growth Rate Random Child Effect	10.89%	10.37%
b_{30}: Spring G1 to Spring G3 Growth Rate Random Child Effect	8.76%	8.11%
b_{40}: Spring G3 to Spring G5 Growth Rate Random Child Effect	1.10%	0.19%
c_{10}: Fall K to Spring K Growth Rate Random School Effect	35.38%	34.08%
e: Level 1 Error	3.45%	2.79%

~ $p < .10$ * $p < .05$ ** $p < .01$

Note: coefficients in this table are unstandardized

Bibliography

Aber, L., Bennett, N., Conley, D., & Li, J. (1997). The effects of poverty on child health and development. *Annual Review of Public Health*, 18, 463-483.

Achenbach, T. M (1966). The classification of children's psychiatric symptoms: a factor-analytic study. *Psychological Monographs,* 80.

Achenbach T. M. (2005). Advancing assessment of children and adolescents: Commentary on evidence-based assessment of child and adolescent disorders. *Journal of Clinical Child and Adolescent Psychology*, 34(3), 541-547.

Achenbach T. M. , Dumenci L., & Rescorla L. A. (2002). Ten-year comparisons of problems and competencies for national samples of youth: Self, parent, and teacher reports. *Journal of Emotional and Behavioral Disorders*, 10(4), 194-203

Achenbach, T. M., McConaughy, S., & Howell, C. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, 101, 213-232.

Ackerman, B. P., Brown, E., & Izard, C. E. (2003). Continuity and change in levels of externalizing behavior in school of children from economically disadvantaged families. *Child Development*, 74(3), 694-709.

Adler, N. E., & Ostrove, J. M. (1999). Socioeconomic status and health: What we know and what we don't. *Anals New York Academy of Sciences*, 896, 3-15.

Agostin, T. M., & Bain, S. K. (1997). Predicting early school success with developmental and social skills screeners. *Psychology in the Schools*, 34(3), 219-228.

Alzaba-Poria, N., Pike, A., & Deater-Deckard, K. (2004). Do risk factors for problem behaviors act in a cumulative manner? The examination of ethnic minority and majority children through an ecological perspective. *Journal of Child Psychology and Psychiatry*, 45(4), 707-718.

Armbruster, P., Gerstein, S. H., Fallon, T. (1997). Bridging the gap between service need and service utilization: a school-based mental health program. *Community Mental Health Journal,* 33(3), 199-211.

Arnold D. H. (1997). Co-occurrence of externalizing behavior problems and emergent academic difficulties in young high-risk boys: A preliminary evaluation of patterns and mechanisms. *Journal of Applied Developmental Psychology*, 18(3), 317-330.

Arnold, D. H., & Doctoroff, G. L. (2003). The early education of socio-economically disadvantaged children. *Annual Review of Psychology*, 54, 517-545.

Baker, J. A. (2006). Contributions of teacher-child relationships to positive school adjustment during elementary school. *Journal of School Psychology*, 44, 211-229.

Baker, J. A., Grant, S., & Morlock, L. (2008). The teacher-student relationship as a developmental context for children with internalizing or externalizing behavior problems. *School Psychology Quarterly*, 23(1), 3-15.

Barnett, W. S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *The Future of Children*, 5(3), 25-50.

Barnett, S. W., Hustedt, J., Robin, K., & Schulman, K (2004). *The state of preschool: 2004 state preschool yearbook*. New Brunswick, NJ: The National Institutes for Early Education Research, Rutgers University.

Barnett, W. S., Jung, K., Yarosz, D. J., Thomas, J., Hornbeck, A., Stechuk R., & Burns, S. (2008). Educational effects of the Tools of the Mind curriculum: A randomized trial. *Early Childhood Research Quarterly,* 23, 299-313.

Bennett, K. J., Brown, K. S., Boyle, M., Racine, Y., & Offord, D. (2003). Does low reading achievement at school entry cause conduct problems? *Social Science and Medicine*, 56, 2443-2448.

Berlin, L. J., Brooks-Gunn, J., McCarton, C., & McCormick, M. C. (1998). The effectiveness of early intervention: Examining risk factors and pathways to enhanced development. *Preventive Medicine*, 27, 238-245.

Bierman, K. L., Domitrovich, C. E., Nix, R. L., Gest, S. D., Welsh, G. A., Greenberg, M. T., ..., & Gill, S. (2008). Promoting academic and social-emotional school readiness: The Head Start REDI program. *Child Development*, 79(6), 1802-1817.

Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *American Psychologist*, 57(2), 111-127.

Blair, C., & Diamond, A. (2008). Biological processes in prevention and intervention: The promotion of self-regulation as a means of preventing school failure. *Development and Psychopathology*, 20, 899-911.

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, 78(2), 647-663.

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2002). *Comprehensive school reform and student achievement.* (Report No. 59). Maryland, MD: Center for Research on the Education of Students Placed At Risk. Retrieved from www.csos.jhu.edu.

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development*.* *Annual Review of Psychology*, 3, 371-399.

Brame, B., Nagin, D.S., & Tremblay, R.E. (2001). Developmental trajectories of physical aggression from school entry to late adolescence. *Journal of Child Psychology and Psychiatry*, *42*, 503–512.

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist*, 32(7), 513-531.

Bronfenbrenner, U. (1979). Contexts of child rearing: Problems and prospects. *American Psychologist*, 34(10), 844-850.

Bronfenbrenner, U., & Morris, P. A. (1998). The ecology of developmental processes. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (5th ed., pp. 993-1029). New York, NY: John Wiley & Sons.

Bronson, M. B. (2000). *Self-regulation in early childhood: Nature and nurture*. New York, NY: Guilford.

Brooks-Gunn J., & Duncan G. J. (1997). The effects of poverty on children. *The Future of Children*, 7, 55-71.

Brooks-Gunn, J., Duncan, G. J., & Aber, J. L. (1997). *Neighborhood poverty: Context and consequences for children.* Vol. 1. New York, NY: Russell Sage Foundation.

Brooks-Gunn, J., & Markman, L. B. (2005). The Contribution of Parenting to Ethnic and Racial Gaps in School Readiness. *Future of Children*, 1, 139-168.

Brown, J. D., Wissow, L. W., Godamski, A., Zachary, C., & Bartlett, E. (2006). Parent and teacher mental health ratings of children using primary-care services: Interrater agreement and implications for mental health screening. *Ambulatory Pediatrics*, 6, 347-351.

Bub, K. L., McCartney, K., & Willett, J. B. (2007). Behavior problem trajectories and first-grade cognitive ability and achievement skills: A latent growth curve analysis. *Journal of Educational Psychology*, 99(3), 653-670.

Burchinal, M. R., Peisner-Feinberg, E., Pianta, R., & Howes, C. (2002). Development of academic skills from preschool through second grade: Family and classroom predictors of developmental trajectories. *Journal of School Psychology*, 40, 415-436.

Burrage, M. S., Ponitz, C. C., McCready, E. A., Shah, P. Sims, B. C., Jewkes, A. M., & Morrison, F. J. (2008). Age- and schooling-related effects on executive functions in young children: A natural experiment. *Child Neuropsychology*, 14(6), 510-524.

Buyse, E., Verschueren, K., Doumen, S., Van Damme, J., & Maes, F. (2008). Classroom problem behavior and teacher-child relationships in kindergarten: The moderating role of classroom climate. *Journal of School Psychology*, 46, 367-391.

Cameron, C. E., Connor, C. M., & Morrison, F. J. (2005). Effects of variation in teacher organization on classroom functioning. *Journal of School Psychology*, 43(1), 61-85.

Campbell, F. A., Ramey, C. T., Pungello, E., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian Project. *Applied Developmental Science*, 6, 42-57.

Campbell, S. B., Shaw, D. D., & Gilliom, M. (2000). Early externalizing behavior problems: Toddlers and preschoolers at risk for later maladjustment. *Development and Psychopathology*, 12, 467-488.

Case, A., Fertig, A., & Paxson, C. (2005). The lasting impact of childhood health and circumstance. *Journal of Health and Economics*, 24, 365-389.

Case, A., Lubotsky, D., & Paxson, C. (2002). Economic status and health in childhood: The origins of the gradient. *The American Economic Review*, 92(5), 1308-1334.

Case, A., & Paxson, C. (2006). Children's health and social mobility. *Future of Children*, 16(2), 151-173.

Casey, P. H., Whiteside-Mansell, L., Barrett, K., Bradley, R. H., & Gargus, R. (2006). Impact of prenatal and/or postnatal growth problems in low birth weigth preterm infants on school-age outcomes: an 8-year longitudinal evaluation. *Pediatrics*, 118(3), 1078-1086.

Caspi, A. (2000). The child is father of the man: personality continuities from childhood to adulthood. *Journal of Personality and Social Psychology*, 78, 158-172

Caughy, M. O. (1996). Health and environmental effect on the academic readiness of school-age children. *Developmental Psychology*, 52(3), 515-522.

Claessens, A., Duncan, G., & Engel, M. (2008). Kindergarten skills and fifth-grade achievement: Evidence from the ECLS-K. *Economics of Education Review*, 28(4), 415-427.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edition. New Jersey: Lawrence Erlbaum Associates.

Cole, P. M., & Deater-Deckard, K. (2009). Emotion regulation, risk, and psychopathology. *Journal of Child Psychology and Psychiatry*, 50(11), 1327-1330.

Cole, P. M., Martin, S. E., & Dennis, T. A. (2004). Emotion regulation as a scientific construct: Methodological challenges and directions for child development research. *Child Development*, 75(2), 317-333.

Coleman, J. S., Campbell, E., Hobson, C., McPartland, J., Mood, F., Weinfeld, F., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.

Colman, R. A., Hardy, S. A., Albert, M., Raffaneli, M., & Crocket, L. (2006). Early predictors of self-regulation in middle childhood. *Infant and Child Development*, 15, 421-437.

Conduct Problems Prevention Research Group (CCPRG). (1999). Initial impact of the Fast Track prevention trial for conduct problems I: The high-risk sample. *Journal of Consulting and Clinical Psychology,* 67, 631-647.

Connor, C. M., Jakobsons, L. J., Crowe, E., & Meadows, J. (2009). Instruction, differentiation, and student engagement in Reading First classrooms. *Elementary School Journal*, 109, 221-250.

Cook, M. D., & Evans, W. N. (2000). Families or schools? Explaining the convergence in White and Black academic performance. *Journal of Labor Economics*, 18, 729–754.

Costello, E. J., Angold, A., Burns, B. J., Stangl, D. K., Tweed, D. L., Erkanli, A., & Worthman, C. M. (1996). The great smoky mountains study of youth: goals, design methods and prevalence of DSM-III-R disorders. *Archives of General Psychiatry*, 53, 1129-1136.

Cunha, F., & Heckman, J. J. (2007). The technology of skill formation. *AEA Papers and Proceedings*, 97(2), 31-47.

Cunha, F., & Heckman, J. J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *The Journal of Human Resources*, 43(4), 738-782.

Cunha, F., Heckman, J. J., Lochner, L. J., & Masterov, D. V. (2006). Interpreting the evidence of life cycle skill formation. In E. A. Hanushek & F. Welch (Eds.), *Handbook of the Economics of Education*, pp. 697-812. Amsterdam: North-Holland.

Currie, J. (2005). Health disparities and gaps in school readiness. *Future of Children*, 15(1), 117-138

Currie, J., & Stabile, M. (2004). Socioeconomic status and health: Why is the relationship stronger for older children? *American Economic Review*, 93(5), 1813-23.

Cutler D. M., & Lleras-Muney A. (2006). Education and Health: Evaluating Theories and Evidence. 2006. NBER working paper No. 12352.

Dauber, S. L., & Epstein, J. L. (1993). Parents' attitudes and practices of involvement in inner-city elementary and middle schools. In N. F. Chavkin (Ed.), *Families and schools in a pluralistic society* (pp. 53-72). Albany, NY State University of New York Press.

De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin*, 131, 483-509.

Denham, S. A., Blair, K. A., DeMulder, E., Levitas, J., Sawyer, K., & Auerbach-Major, S. (2003). Preschool emotional competence: Pathway to social competence? *Child Development*, 74(1), 238-256.

Diamond, A. (2006). The early development of executive functions. In E. Bialystok & F. Craik (Eds.), *Lifespan cognition: Mechanisms of change* (pp. 70-95). New York, NY: Oxford University Press.

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science*, 318, 1387-1388

Doctoroff G. L., Greer J. A., & Arnold D. H. (2006). The relationship between social behavior and emergent literacy among preschool boys and girls. *Journal of Applied Developmental Psychology*, 27(1), 1-13.

Dobbs, J., Doctoroff, G. L., Fisher, P. H., & Arnold, D. H. (2006). The association between preschool children's socio-emotional functioning and their mathematical skills. *Journal of Applied Developmental Psychology*, 27(2), 97-108.

Domitrovich, C. E., Cortes, R. C., & Greenberg, M. T. (2007). Improving young children's social and emotional competence: A randomized trial of the preschool "PATHS" curriculum. *The Journal of Primary Prevention*, 28, 67-91.

Dowsett, S. M., & Livesey, D. J. (2000). The development of inhibitory control in preschool children: Effects of "Executive Skills" training. *Developmental Psychobiology*, 36(2), 161-174.

Duncan, G.S., & Brooks-Gun, J. (2000). Family poverty, welfare reform, and child development. *Child Development*, 71(1), 188-196.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., …, & Duckworth, K. (2007). School readiness and later achievement. *Developmental Psychology*, 43, 1428-1446.

Eisenberg, N., Champion, C., & Ma, Y. (2004). Emotion-related regulation: An emerging construct. *Merrill-Palmer Quarterly*, 50, 236-259.

Essex, M. J., Kraemer, H. C., Armstrom, J. H., Boyce, T., Goldsmith, H. H., Kelin, M. H., … & Kupfer, D. (2006). Exploring risk factors for the emergence of children's mental health problems. *Archives of General Psychiatry*, 65, 1296-1256.

Evans, G. W. (2004). The environment of childhood poverty. *American Psychologist*, 59(2), 77-92.

Fantuzzo, J. W., Rouse, H. L., McDermott, P. A., Sekino, Y., Childs, S., & Weiss, A. (2005). Early childhood experiences and kindergarten success: A population-based study of a large urban setting. *School Psychology Review*, 34(4), 571-588.

Fendrich, M., Johnson, T., Wislar, J. S., & Nageotte, C. (1999). Accuracy of parent mental health service reporting: Results from a reverse record-check study. *Journal of American Academy of Child & Adolescent Psychiatry*, 38(2), 147-155.

Feuerstein, A. (2000). School characteristics and parent involvement: Influences on participation in children's schools. *The Journal of Educational Research*, 94, 29-38.

Firth, D. & Bennett, K. E. (1998). Robust models in probability sampling. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 60(1), 3-21.

Fiscella, K., & Kitzman, H. (2009). Disparities in academic achievement and health: The intersection of child education and health policy. *Pediatrics*, 123, 1073-1080.

Fischer, K. W., Bullock, D. H., Rotenberg, E. J., & Raya, P. (1993). The dynamics of competence: How context contributes directly to skill. In R. H. Wozniak, K. W. Fischer, et al (Eds.) *Development in context: Acting and thinking in specific environments*. The Jean Piaget symposium series. (pp. 93-117). Hillsdayle, NJ, USA: Lawrence Erlbaum Associates, Inc. xvi.

Flanagan, K. S., Bierman, K. L., & Kam, C. M. (2003). Identifying at-risk children at school entry: The usefulness of multibehavioral problem profiles. *Journal of Clinical Child & Adolescent Psychology*, 32(3), 396-407.

Foster, E. M. (2002). How economists think about family resources and child development. *Child Development*, 73, 1904-1914.

Foster, S., Rollefson, M., Doksum, T., Noonan, D., Robinson, G., & Teich, J. (2005). *School Mental Health Services in the United States, 2002–2003*. DHHS Pub. No. (SMA) 05-4068. Rockville, MD: Center for Mental Health Services, Substance Abuse and Mental Health Services Administration.

Gall, G., Pagano, M. E., Desmond, M. S., Perrin, J. M., & Murphy, J. M. (2000). Utility of psychosocial screening at a school-based health center. *Journal of School Health*, 70, 292-298.

Garrett, P., Ng'andu, N., & Ferron, J. (1994). Poverty experiences of young children and the quality of their home environments. *Child Development*, 64, 331–345.

Gathercole, S. E. & Baddeley, A. D. (1993). Phonological working memory: A critical building block for reading development and vocabulary acquisition? *European Journal of Psychology of Education,* 8(3), 259-272.

Gathercole, S. E., & Pickering, S. J. (2000). Assessment of working memory in six- and seven-year-old children*.* *Journal of Educational Psychology*, 92(2), 377-390.

Geierstander, S. P., Amaral, G., Mansour, M., & Walters, S. R. (2004). School-based health centers and academic performance: Research, challenges, and recommendations. *The Journal of School Health*. 74(9), 347-352.

Gershoff, E. T. (2003). *Low income and the development of America's kindergartners. Living at the edge research brief*. New York, NY: National Center for Children in Poverty.

Glascoe, F. P. (2000). Evidence-based approach to developmental and behavioral surveillance using parents' concerns. *Child: Care, Health, and Development*, 26(2), 137-149.

Glewwe, P., Jaboy, H. G., & King, E. M. (2001). Early childhood nutrition and academic achievement: A longitudinal analysis. *Journal of Public Economics*, 81, 345-368.

Goldman, N. (2001). Social inequalities in health: disentangling the underlying mechanisms. In Weinstein, M., Hermalin, A. (Eds.), *Strengthening the dialogue between epidemiology and demography*. Annals of the New York Academy of Sciences.

Graziano, P. A., Reavis, R. D., Keane, S. P., & Calkins, S. D. (2007). The role of emotion regulation in children's early academic success. *Journal of School Psychology*, 45, 3-19.

Gresham, F. M., & Elliott, S.N. (1990). *Social Skills Rating System Manual.* Circle Pines: American Guidance Service.

Gutman, L. M., Sameroff, A. J., & Cole, R. (2003). Academic growth curve trajectories from 1st grade to 12th grade: effects of multiple social risk factors and preschool child factors. *Developmental Psychology*, 39(4), 777-790.

Hair, E., Halle, T., Terry-Humen, E., Lavelle, B., & Calkins, J. (2006). Children's school readiness in the ECLS-K: predictions to academic, health, and social outcomes in first grade. *Early Childhood Research Quarterly*, 21, 431-44.

Hamre, B. K., & Pianta, R. C. (2001). Early teacher-child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development*, 72, 625-638.

Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development*, 76(5), 949-967.

Hanson, T.L., Austin, G. A., & Lee-Bayha, J. (2004). Ensuring that no child is left behind: How are student health risks and resilience related to the academic progress of schools? WESTED report, Retrieved from http://www.wested.org/hks

Hansen, B. B., & Bowers, J. (2008). Covariate balance in simple, stratified, and clustered comparative studies. *Statisitical Science*, 23(2), 219-236.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, 312, 1900-1902.

Heckman, J. J. (2007). The economics, technology, and neuroscience of human capability formation. *Proceedings of the National Academy of Sciences*, 104(33), 13250-13255.

Hedges, L.V., Laine, R.D., & Greenwald, R. (1994). Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher*, 23(3), 5-14.

Henderson, A., & Berla, N. (1994). *A new generation of evidence: The family is critical to student achievement.* Washington, DC: Center for Law and Education.

Henricsson, L., & Rydell, A.M. (2006). Children with behavior problems: The influence of social competence and social relations on problem stability, school achievement and peer acceptance across the first six years of school. *Infant and Child Development*, 15, 347-366.

Hinshaw, S. P., (1992a). Academic underachievement, attention deficits, and aggression: comorbidity and implications for intervention. *Journal of Consulting and Clinical Psychology*, 60(6), 893-903.

Hinshaw, S. P., (1992b). Externalizing behavior problems and academic underachievement in childhood and adolescence: causal relationships and underlying mechanisms. *Psychological Bulletin*, 111, 127-155.

Hinshaw, S. P. (2002). Process, mechanism, and explanation related to externalizing behavior in developmental psychopathology. *Journal of Abnormal Child Psychology,* 30(5), 431-446.

Hinshaw, S. P., Han, S., Erhardt, D., & Huber-Dressler, A. (1992). Parent ratings, teacher ratings, and direct observations of preschool children: Differential convergence for externalizing vs. internalizing behavior patterns. *Journal of Clinical Child Psychology*, 21, 143-150.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(369), 945-960.

Hong, G., & Raudenbush, S. W. (2008). Causal inference for time-varying instructional treatments. *Journal of Educational and Behavioral Statistics,* 33(3), 333-362.

Howse, R., Calkins, S., Anastopoulos, A., Keane, S., & Shelton, T. (2003). Regulatory contributors to children's academic achievement. *Early Education and Development*, 14, 101-119.

Howse, R. B., Lange, G., Farran, D. C., & Boyles, C. D. (2003). Motivation and self-regulation as predictors of achievement in economically disadvantaged young children. *The Journal of Experimental Education*, 71, 151–174.

Hughes, J. N., Luo, W., Kwork, O., & Lloyd, L. K. (2008). Teacher-student support, effortful engagement, and achievement: a 3-year longitudinal study. *Journal of Educational Psychology*, 100(1), 1-14.

Hughes, C., White, A., Sharpen, J., & Dunn, J. (2000). Antisocial, angry, and unsympathetic: "Hard-to-manage" preschoolers' peer problems and possible cognitive influences. *Journal of Child Psychology & Psychiatry & Allied Disciplines*, 41(2), 169-179.

Huston, A. C., & Ripke, M. N. (2006). Middle Childhood: Contexts of development. In A. C. Huston, & M. N. Ripke. *Developmental contexts in middle childhood: Bridges to adolescence and adulthood* (pp. 217-236). Cambridge, MA: Cambridge University Press.

Jackson, A. (2003). The effects of family and neighborhood characteristics on the behavioral and cognitive development of poor black children: A longitudinal study. *American Journal of Community Psychology*, 32(1/2), 175-186.

Jacob, R., & Parkinson, J. (working paper). Can executive functions be trained? A meta-analysis of executive function training programs and interventions.

Jamner, M. S., Spruijt-Metz, D., Bassin, S., & Cooper, D. M. (2004). A controlled evaluation of a school-based intervention to promote physical activity among sedentary adolescent females: Project FAB. *Journal of Adolescent Health*, 34(4), 278-289.

Jencks, C., & Phillips, M. (Eds.) (1998). *The Black-White test score gap.* Washington, DC: Brookings Institution Press.

Jensen, P. S., Rubio-Stipec, M., Canino, G., Bird, H. R., Dulcan, M. K., Schwab-Stone, M. E., & Lahey, B. B. (1999). Parent and child contributions to diagnosis of mental disorder: Are both informants always necessary?. *Journal of the American Academy of Child and Adolescent Psychiatry,* 38, 1569-1579.

Jimerson, S., Egeland, B., & Teo, A. (1999). A longitudinal study of achievement trajectories: factors associated with change. *Journal of Educational Psychology*, 91, 116-126.

Johnson, W., McGue, M., & Iacono, W. G. (2006). Genetic and environmental influences on academic achievement trajectories during adolescence. *Developmental Psychology*, 42(3), 514-532.

Jones, S. M., Brown, J. L., Hoglund, W., & Aber, J. L. (under review). Impacts of an integrated school-based social-emotional learning and literacy intervention on third-grade social-emotional and academic outcomes, *Journal of Consulting and Clinical Psychology.*

Kagan, S. L., Moore, E., & Bredekamp, S. (Eds.). (1995). *Reconsidering Children's Early Learning and Development: Toward Shared Beliefs and Vocabulary.* Washington, DC: National Education Goals Panel.

Kail, R. V. (2003). Information processing and memory. In M. H. Bornstein, L. Davidson, C. L. M. Keyes, & K. A. Moore (Eds.), *Crosscurrents in contemporary psychology* (pp. 269-279). Mahwah, NJ: Lawrence Erlbaum Publishers.

Kam, C. M., Greenberg, M. T., & Walls, C. T. (2003). Examining the role of implementation quality in school-based prevention using the PATHS curriculum. *Prevention Science,* 4, 55-63.

Kaplan, D. W., Calonge, B. N., Guernsey, B. P., & Hanrahan, M. B. (1998). Managed care and school-based health centers: Use of health services. *Archives of Pediatric Adolescent Medicine*, 152, 25-33.

Keenan, K., Shaw, D., Delliquadri, E., Giovannelli, J., & Walsh, B. (1998). Evidence for the continuity of early problem behaviors: application of a developmental model. *Journal of Abnormal Child Psychology*, 26, 441-454.

Keiley, M. K., Bates, J. E., Dodge, K. A., & Pettit, G. S. (2000). A cross-domain growth analysis: Externalizing and internalizing behaviors during 8 years of childhood. *Journal of Abnormal Child Psychology*, 28, 161-179.

Kerr, A., & Zelazo, P. D. (2004). Development of "hot" executive function: The Children's Gambling Task. *Brain and Cognition*, 55, 148-157.

Kisker, E. E., & Brown, R. S. (1996). Do school-based health centers improve adolescents' ccess to health care, health status, and risk-taking behavior? *Journal of Adolescent Health*, 18, 335-343.

Klebanov, P. K., Brooks-Gunn, J., McCarton, C., & McCormick, M. C. (1998). The contribution of neighborhood and family income to developmental test scores over the first three years of life. *Child Development*, 69(5), 1420-1436.

Kochanska, G., & Knaack, A. (2003). Effortful control as a personality characteristic of young children: Antecedents, correlates, and consequences. *Journal of Personality*, 71, 1087-1112.

Kochanska, G., Murray, K. T., & Harlan, E. T. (2000). Effortful control in early childhood: Continuity and change, antecedents, and implications for social development. *Developmental Psychology*, 36, 220-232.

Konstantopoulos, S. (2006). Trends of school effects on student achievement: Evidence from NLS:72, HSB:82, and NELS:92. *Teachers College Record*, 108(12), 2550-2581.

Kowaleski-Jones, L., & Duncan, G. J. (1999). The structure of achievement and behavior across middle childhood. *Child Development*, 4, 930-943.

Kraemer, H. C., Measelle, J. R., Ablow, J. C., Essex, M.J., Boyce, W.T., & Kupfer, D.J. (2003). A new approach to multiple informants: Mixing and matching context and perspectives. American. *Journal of Psychiatry*, 160, 1566-1577.

Ladd, G. W., & Burgess, K. B. (1999). Charting the relationship trajectories of aggressive, withdrawn, and aggressive/withdrawn children during early grade school. *Child Development*, 70, 910-929.

Ladd, G. W., Herald, S. L., & Kochel, K. P. (2006). School readiness: Are there social prerequisites? *Early Education and Development*, 17, 115-150.

Lahey, B. B., Loeber, R. Hart, E. L., Frick, P. J., Applegate, B., Zhang, Q., …, & Russo, M. F. (1995). Four-year longitudinal study of conduct disorder in boys: patterns and predictors of persistence. *Journal of Abnormal Psychology*, 104, 83-93.

Lahey, B. B., McBurnett, K., & Loeber, R. (2000). Are attention-deficit/hyperactivity disorder and oppositional disorder developmental precursors to conduct disorder? In A. J. Sameroff, M. Lewis, & S. M. Miller (Eds.), *Handbook of developmental psychopathology* (pp. 431-446). New York, NY: Kluwer Academic/Plenum Publishers.

Leerkes, E. Paradise, M. J., O'Brien, M., Calkins, S. D., & Garrett, L. (2008). Emotion and cognition processes in preschool children. *Merrill-Palmer Quarterly*, 54, 102-124.

Levitt, J. M., Saka N., Romanelli L. H., & Hoagwood, K. (2007). Early identification of mental health problems in schools: The status of instrumentation. *Journal of School Psychology*, 45(2), 163-191.

Loeber, R. Farrington, D. P., Stauthamer-Loeber, M., & Van Kammen, W. B. (1998). *Antisocial behavior and mental health problems.* Mahway, NJL Erlbaum.

Lwebuga-Mukasa, J., & Dunn-Georgiou, E. (2002). A school-based asthma intervention program in the Buffalo, New York, schools. *The Journal of School Health,* 72(1), 27-32.

Malecki, C.K., & Elliott, S.N. (2002). Children's social behaviors as predictors of academic achievement: A longitudinal analysis. *School Psychology Quarterly*, 17(1), 1–23.

Mashburn, A. J., Hamre, B. K., Downer, J. T., & Pianta, R. C. (2006). Teacher and classroom characteristics associated with teachers' ratings of prekindergartners' relationships and behaviors. *Journal of Psychoeducational Assessment*, 24(4), 367-380.

Masten, A. S., & Coatsworth, J. D. (1998). The development of competence in favorable and unfavorable environments. *American Psychologist*, 53(2), 205-220.

McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly*, 21, 471-490.

McClelland, M. M., Cameron, C. E., Connor, C. D., Farris, C. L., Jewkes, A. M., & Morrison, F. J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology*, 43(4), 947-959.

McClelland, M. M., Cameron, C., Wanless, S., & Murray, A. (2007). Executive function, self-regulation, and social-emotional competence: Links to school readiness. In O. N. Saracho & B. Spodek (Eds.), *Contemporary perspectives on research in social learning in early childhood education*. (pp. 83-107). Charlotte, NC: Information Age.

McClelland, M. M., & Morrison, F. J. (2003). The emergence of learning-related social skills in preschool children. *Early Childhood Research Quarterly*, 18, 206-224.

McClelland, M. M., Morrison, F. J., & Holmes, D. L. (2000). Children at risk for early academic problems: The role of learning-related social skills. *Early Childhood Research Quarterly*, 15(3), 307-329.

McCrea, S. M., Mueller, J. H., & Parrila, R. K. (1999). Quantitative analyses of schooling effects on executive function in young children. *Child Neuropsychology*, 5(4): 242-250.

McGee, R., Prior, M., Williams, S., Smart, D., & Sanson, A. (2002). The long-term significance of teacher-rated hyperactivity and reading ability in childhood: Findings from two longitudinal studies. *Journal of Child Psychology and Psychiatry*, 43, 1004-1017.

McLeod, J. D., & Fettes, D. L. (2007). Trajectories of failure: The educational careers of children with mental health problems. *American Journal of Sociology*, 113(3), 653-701.

McLeod, J. D., & Kaiser, K. (2004). Childhood Emotional and Behavioral Problems and Educational Attainment. *American Sociological Review*, 69(5), 636-658.

McLoyd, V. (1998). Socioeconomic disadvantage and child development. *American Psychologist*, 53, 185-204.

Mercer, S. H., & DeRosier, M. E. (2008). Teacher preference, peer rejection, and student aggression: A prospective study of transactional influence and independent contributions to emotional adjustment and grades. *Journal of School Psychology*, 46, 661-685.

Merrell, K.W. (2003). *Behavioral, social, and emotional assessment of children and adolescents* (2nd Ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Meyers, J L., & Beretvas, S. N. (2006). The impact of inappropriate modeling of cross-classified data structures. *Multivariate Behavioral Research*, 41(4), 473-497.

Miles, S. B., & Stipek, D. (2006). Contemporaneous and longitudinal associations between social behavior and literacy achievement in a sample of low-income elementary school children. *Child Development*, *77*(1), 103-117.

Mitchell-Herzfeld, S., Izzo, C., Greene, R., Lee, E., & Lowenfels, A. (2005). *Evaluation of Healthy Families New York (HFNY): First year program impacts.* Report to the Governor and Legislature: New York State Office of Children and Family Services.

Moffit, T. E., Caspi, A., Harrington, H., & Milne, B. J. (2002). Males on the life-course-persistent and adolescent-limited antisocial pathways: Follow-up at age 26 years. *Development and Psychopathology*, 14, 179-207.

Morgan, P. L., Farkas, G., Tufis, P. A., & Sperling, R. A. (2008). Are reading and behavior problems risk factors for each other? *Journal of Learning Disabilities*, 41, 417-436

Morrison, F. J., & Connor, C. M. (2002). Understanding schooling effects on early literacy: A working research strategy. *Journal of School Psychology*, 40(6), 493-500.

Morrison, F. J., Ponitz, C. C., & McClelland, M. M. (2009). Self-regulation and academic achievement in the transition to school. In S. D. Calkins & M. A. Bell (Eds.), *Child development at the intersection of emotion and cognition*. Washington: APA Books.

Murray, K. T., & Kochanska, G. (2002). Effortful control: Relation to externalizing and internalizing behaviors and factor structure. *Journal of Abnormal Child Psychology*, 30, 503-514.

Najman, J. M., Aird, R., Bor, W., O-Callaghan, M., Williams, G. M., & Shuttlewood, G. J. (2004). The generational transmission of socioeconomic inequalities in child cognitive development and emotional health. *Social Science and Medicine*, 58, 1147-1158.

National Research Council and Institute of Medicine (NRCIM). (2004). *Children's health, the nation's wealth: Assessing and improving child health.* Washington, DC: The National Academies Press.

National Center for Education Statistics (NCES). (2009). *National Assessment of Educational Progress Grade 4 National Results.* Retrieved from http://nationsreportcard.gov/reading_2009/nat_g4.asp

National Institute of Child Health and Human Development (NICHD) Early Child Care Research Network (ECCRN). (2004). Does Class Size in First Grade Relate to Children's Academic and Social Performance or Observed Classroom Processes? *Developmental Psychology*, 40(5), 651-664.

Neumark-Sztainer, D., Martin, S. L., & Story, M. (2000). School-based programs for obesity prevention: What do adolescents recommend? *American Journal of*

Health Promotion, 14(4), 232-235.

Neumark-Sztainer, D., & Story, M. (1997). Recommendations from overweight youth regarding school-based weight control programs. *Journal of School Health*, 67(10), 428-433.

Newacheck, P., Jameson, W. J., & Halfon, N. (1994). Health status and income: The impact of poverty on child health. *The Journal of School Health*, 64(6), 229-233.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). Do minorities experience larger lasting benefits from small classes? Evidence from a five-year follow-up of the Tennessee class size experiment. *Journal of Educational Research,* 98, 94-100.

Offord D. R., Boyle, M. H., Racine, Y., Szatmari, P., Fleming J. E., Sanford M., & Lipman E. L., (1996). Integrating assessment data from multiple informants. *Journal of the American Academy of Child and Adolescent Psychiatry*, 35, 1078-1085.

Olds, D. L., Robins, J., O'Brien, R., Luckey, D., Pettitt, L. M., Henders, C. R., … & Talmi, A. (2002). Home visiting by paraprofessionals and by nurses: A randomized, controlled trial. *Pediatric*s, 110(3), 486-496.

Ollendick, T. H., & Hersen, M. (1993). *Handbook of child and adolescent assessment*. Boston, MA: Allyn and Bacon.

Palloni, A. (2006). Reproducing inequalities: luck, wallets, and the enduring effects of childhood health. *Demography*, 43(4), 587-615.

Parkinson, J., & Rowan, B. (2008). Poverty, literacy achievement, and education reform. In S. B. Neuman (Ed.) *Educating the other America: Top experts tackle poverty, literacy, and achievement in our schools*, Baltimore, MD: Paul H Brookes Pub.

Pastor, P. N., Reuben, C. A., & Falkenstern, A. (2004). Parental reports of emotional or behavioral difficulties and mental health service use among U.S. school-age children. In R. W. Manderscheid, J. T. Berry (Eds.) *Mental health, United States, 2004*. (pp. 189-199). Rockville, Maryland.

Patrick, H. (1997). Social self-regulation: Exploring the relations between children's social relationships, academic self-regulation and school performance, *Educational Psychologist*, 32(4), 209-220.

Patterson, G. R., Forgatch, M. S., Yoerger, K. L., & Stoolmiller, M. (1998). Variables that initiate and maintain an early-onset trajectory for juvenile offending. *Development and Psychopathology*, 10, 531-547.

Payton, J., Weissberg, R. P., Durlak, J. A., Dymnicki, A. B., Taylor, R. D., Schellinger, K. B., & Pachan, M. (2008). *The positive impact of social and emotional learning for kindergarten to eight-grade students: Findings from three scientific reviews*. Chicago, IL: Collaborative for Academic, Social, and Emotional Learning.

Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., & Yazejian, N. (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development*, 72(5), 1534-1553.

Pianta, R. C. (2006). Schools, schooling, and developmental psychopathology. In D. Cicchetti (Ed.), *Handbook of developmental psychopathology* (Vol. 2), New York, NY: John Wiley & Sons.

Pollack, J., Atkins-Burnett, S., Rock, D., & Weiss, M. (2005). *Early childhood longitudinal study, kindergarten class of 1998–99 (ECLS-K), psychometric report for the third grade* (NCES 2005–062). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Ponitz, C. C., McClelland, M. M., Jewkes, A. M., Connor, C. M., Farris, C. L., & Morrison, F. J. (2008). Touch your toes! Developing a direct measure of behavioral regulation in early childhood. *Early Childhood Research Quarterly,* 23, 141–158.

Ponitz, C. C., McClelland, M. M., Matthews, J. S., & Morrison, F. J. (2009). A structured observation of behavioral self-regulation and its contribution to kindergarten outcomes. *Developmental Psychology*, 45(3), 605-619.

Ponitz, C. C., Rimm-Kaufman, S. E., Brock, L. L., & Nathanson, L. (2009). Early adjustment, gender differences, and classroom organizational climate in first grade. *The Elementary School Journal*, 110(2), 142-162.

Powell, C. A., Walker, S. P., Chang, S. M., & Grantham-McGregor, S. M. (1998). Nutrition and education: A randomized trial of the effects of breakfast in rural primary school children. *American Journal of Clinical Nutrition*, 68, 873-879.

Rabiner, D., Coie, J. D., & The Conduct Problems Prevention Research Group (CPPRG). (2000). Early attention problems and children's reading achievement: A longitudinal investigation. *Journal of the American Academy of Child and Adolescent Psychiatry*, 39, 859-867.

Rapport, M. D., Denney, C. B., Chung, K., & Hustace, K. (2001). Internalizing behavior problems and scholastic achievement in children: Cognitive and behavioral pathways as mediators of outcome. *Journal of Clinical Child Psychology*, 30(4), 536-551.

Rathbun, A., & West, J. (2004). *From kindergarten through third grade: children's beginning school experiences* (NCES 2004–007). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Raudenbush, S. W. (2001). Comparing personal trajectories and drawing causal inferences from longitudinal data. *Annual Review of Psychology*, 52, 501-525.

Raudenbush, S. (2004). What are value-added models estimating and what does this imply for statistical practice? *Journal of Educational and Behavioral Statistics*, 29(1), 121-129.

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

Raver, C. C. (2004). Placing emotional self-regulation in sociocultural and socioeconomic contexts. *Child Development*, 75(2), 346-353.

Raver, C. C., Garner, P., & Smith-Donald, R. (2007). The roles of emotion regulation and emotion knowledge for children's academic readiness: Are the links causal? In R. C. Pianta, M. J. Cox, & K. L. Snow (Eds.), *School readiness and the transition to kindergarten in the era of accountability* (pp. 121–147). Baltimore, MD: Brookes Publishing.

Raver, C. C., Gershoff, E. T., & Aber, J. L. (2007). Testing equivalence of mediating models of income, parenting, and school readiness for White, Black, and Hispanic children in a national sample. *Child Development*, 78(1), 96-115.

Raver, C. C., Jones, S. M., LiGrining, C., Zhai, F., Metzger, M. W., & Solomon, B. (2009). Targeting children's behavior problems in preschool classrooms: A cluster-randomized controlled trial. *Journal of Consulting and Clinical Psychology*, 77(2), 302-316.

Raver, C. C., & Zigler, E. F. (1997). Social competence: An untapped dimension of Head Start's success. *Early Childhood Research Quarterly,* 12, 363-385.

Reardon, S. (2008). Differential growth in the Black-White achievement gap during elementary school among initially high-and low-scoring students. Working Paper.

Reichman, N. E. (2005). Low birth weight and school readiness. *Future of Children*, 15(1), 91-116.

Repetti, R. L., Taylor, S. E., & Seeman, T. E. (2002). Risky families: Family social environments and the mental and physical health of offspring. *Psychological Bulletin*, 128(2), 330-366

Richters, J. (1992). Depressed mothers as informants about their children: A critical review of the evidence for distortion. *Psychological Bulletin*, 112, 485-499.

Rimm-Kaufman, S. E., Pianta, R. C., & Cox, M. J., (2000). Teachers' judgments of problems in the transition to kindergarten. *Early Childhood Research Quarterly*, 15, 147-166.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, 73, 417-458.

Roeser, R. W., & Eccles, J. S. (1998). Adolescents' perceptions of middle school: relation to longitudinal changes in academic and psychological adjustment. *Journal of Research on Adolescence*, 8, 123-158.

Roeser, R. W., Eccles, J. S., & Strobel, K. R. (1998). Linking the study of schooling and mental health: selected issues and empirical illustrations at the level of the individual. *Educational Psychologist*, 33(4), 153-176.

Rosenbaum, P. R. (2002). *Observational Studies*. Cambridge, MA: Cambridge University Press.

Rothbart, M. K., & Bates, J. E. (1998). Temperament. In N. Eisenberg (Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (5th ed.), New York, NY: Wiley.

Rothbart, M. K., & Bates, J. E. (2006). Temperament. In N. Eisenberg (Ed.), *Handbook of child psychology* (6th ed., Vol. 3, pp. 99-106). Hoboken, NJ: Wiley.

Rothbart, M. K., & Posner, M. I. (2005). Genes and experience in the development of executive attention and effortful control. *New Directions for Child and Adolescent Development*, 109, 101-108.

Rothstein, R. (2004). The achievement gap: a broader picture. *Educational Leadership*, 62(3), 40-43.

Rowan, B., Correnti, R., & Miller, R. (2002). What large-scale survey research tells us about teacher effects on student achievement: Insights from the Prospects study of elementary schools. *Teachers College Record*, 104, 1525-1567.

Rowe, K. J., & Rowe, K. S. (1999). Investigating the relationship between students' attentive–inattentive behaviors in the classroom and their literacy progress: Chapter 4 Model specifications, results and comments: fitting multi–level, non-recursive structural equation models. *International Journal of Educational Research*, 31(1), 49-64.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York, NY: John Wiley & Sons, Inc.

Rueda, M. R., Posner, M. I., & Rothbart, M. K. (2005). The development of executive attention: Contributions to the emergence of self-regulation. *Developmental Neuropsychology*, 28(2), 573-594.

Sameroff, A. J. (1986). Environmental context of child development. *The Journal of Pediatrics*, 109(1), 192-200.

Sameroff, A. J. (2000). Developmental systems and psychopathology. *Developmental and Psychopathology*, 12, 297-312.

Sameroff, A. J., Fiese, B. H. (2000). Transactional regulation: The developmental ecology of early intervention. In Shonkoff, J. P., & Meisel, S. J. (Eds.), *Handbook of Early Childhood*. (pp.) Cambridge, MA: Cambridge University Press.

Sameroff, A. J., & Mackenzie, M. J. (2003). Research strategies for capturing transactional models of development: The limits of the possible. *Development and Psychopathology*, 15, 613-640.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*, New York, NY: Chapman and Hall.

Schulenberg, J. E., Maggs, J. L., & O'Malley, P. M. (2003). How and why the understanding of developmental continuity and discontinuity is important: The sample case of long-term consequences of adolescent substance use. In J. T. Mortimer & M. J. Shanahan (Eds.) *Handbook of the Life Course*. (pp. 413-436).New York, NY: Kluwer Academic/Plenum Publishers.

Scott-Jones, D. (1995). Parent–child interactions and school achievement. In B. A. Ryan, G. R. Adams, T. P. Gullotta, R. P. Weissberg, & R. L. Hampton (Eds.), *The family–school connection: Vol. 2. Theory, research, and practice* (pp. 75 – 109). Thousand Oaks, CA: Sage.

Shields, A., Dickstein, S., Seifer, R., Giusti, L., Magee, K. D., & Spritz, B. (2001). Emotional competence and early school adjustment: A study of preschoolers at risk. *Early Education & Development*, 12(1), 73-96.

Shonkoff, J. P., & Phillips, D. A. (2000). *From neurons to neighborhoods: The science of early childhood development*, Washington, DC: National Academy Press.

Silver, R. B., Measelle, J. R., Armstrong, J. M., & Essex, M. J. (2005). Trajectories of classroom externalizing behavior: contributions of child characteristics, family characteristics, and the teacher-child relationship during the school transition. *Journal of School Psychology*, 43, 39-60.

Spinrad, T. L., Eisenberg, N., Cumberland, A., Fabes, R. A., Valiente, C., Shepard, S. A., …, & Guthrie, I. K. (2006). Relation of emotion-related regulation to children's social competence: A longitudinal study. *Emotion*, 6(3), 498-510.

Spira, E. G., & Fischel, J. E. (2005). The impact of preschool inattention, hyperactivity, and impulsivity on social and academic development: A review. *Journal of Child Psychology and Psychiatry*, 46(7), 755-773.

Sroufe, L. A. (1996). *Emotional development*. New York, NY: Cambridge University Press.

Stevens, G. D. (2006). Gradients in the health status and developmental risks of young children: The combined influences of multiple social risk factors. *Maternal and Child Health Journal*, 10(2), 187-199.

Strohschein, L. (2005). Household income histories and child mental health trajectories. *Journal of Health and Social Behavior*, 46(4), 359-375.

Symons, C. W., Cinelli, B., James T. C., & Groff, P. (1997). Bridging student health risks and academic achievement through comprehensive school health programs. *The Journal of School Health*. 67(6), 220-227.

Taras, H., Write, S., Brennan, J., Campana, J., & Lofgren, R. (2004). Impact of school nurse case management of students with asthma. *The Journal of School Health*, 74(6), 213-219.

Tate, R. L. (2000). Elaboration of HLM growth modeling results. *Florida Journal of Educational Research*, 40(2), 53-75.

Teo, A., Carlson, E., Mathieu, P. J., Egeland, B., & Sroufe, L. A. (1996). A prospective longitudinal study of psychosocial predictors of achievement. *Journal of School Psychology*, 34, 285-306.

Tourangeau, K., Lê, T., and Nord, C. (2005). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS–K), Fifth-Grade Methodology Report* (NCES 2006–037). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Tremblay, R. E. (2000). The development of aggressive behavior during childhood: What have we learned in the past century? *International Journal of Behavioral Development*, 24, 129- 141.

Trzesniewski, K. H., Moffitt, T. E., Caspi, A., Taylor, A., & Maughan, B. (2006). Revisiting the association between reading achievement and antisocial behavior: New evidence of an environmental explanation from a twin study. *Child Development,* 77(1), 72-88.

U.S. Department of Health and Human Services (US DHHS). (1999). *Mental health: A report of the Surgeon General.* Rockville, MD: U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Mental Health Services, National Institutes of Health, National Institute of Mental Health.

US DHHS. (2001). *Mental health: Culture, race, and ethnicity—A supplement to Mental health: A report of the Surgeon General*. Rockville, MD: U.S. Department of Health and Human Services, Public Health Service, Office of the Surgeon General.

U.S. Department of Education, National Center for Education Statistics. (2006). *ECLS-K Longitudinal Kindergarten-Fifth Grade Public-Use Data File and Electronic Codebook* (CD-ROM). (NCES 2006-035). Washington, DC: Author.

Vazsonyi, A. T., & Huang, L. H. (2010). Where self-control comes from: On the development of self-control and its relationship to deviance over time. *Developmental Psychology*, 46(1), 245-257.

Verhulst, F. C., Achenbach, T. M., van der Ende, J., Role, N., Lambert, M. C., Leung, P. W. L., ..., & Zubrick, S. R. (2003). Comparisons of problems reported by youths from seven countries. *American Journal of Psychiatry*, 160(8), 1479-1485.

Vitaro, F., Brendgen, M., Larose, S., & Tremblay, R. E. (2005). Kindergarten disruptive behaviors, protective factors, and educational achievement by early adulthood. *Journal of Educational Psychology*, 97(4), 617-629.

von Suchodoletz, A., Trommsdorff, G., Heikamp, T., Wieber, F., & Gollwitzer, P. M. (2009). Transition to school: The role of kindergarten children's behavior regulation. *Learning and Individual Differences*, 19, 561-566.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard Press.

Welsh, J. E., & Bierman, K. L. (2001). Social competence. *Encyclopedia of Childhood and Adolescence*. Retrieved from http://findarticles.com/p/articles/mi_g2602/is_0004/ai_2602000487/.

Wentzel, K. R. (1991). Relations between social competence and academic achievement in early adolescence. *Child Development*, 62(5), 1066-1078.

Wentzel, K. R. (1993). Does being good make the grade? Relations between academic and social competence in early adolescence, *Journal of Educational Psychology*, 85, 357-364.

Wilding, J. (2005). Is attention impaired in ADHD? *British Journal of Developmental Psychology*, 23(4), 487-505.

Wilson, R. S. (1985). Risk and resilience in early mental development. *Developmental Psychology*, 21(5), 795-805.

Wilson, D. B., Gottfredson, D. C., & Najaka, S. S. (2001). School-based prevention of problem behaviors: A meta-analysis. *Journal of Quantitative Criminology*, 17, 247-272.

Wilson, H. K., Pianta, R. C., & Stuhlman, M. (2007). Typical classroom experiences in first grade: The role of classroom climate and functional risk in the development of social competencies. *The Elementary School Journal*, 108(2), 81-96.

Xue, Y., Leventhal, T., Brooks-Gunn, J., & Earls, F. J. (2005). Neighborhood residence and mental health problem of 5- to 11-year-olds. *Archives of General Psychiatry*, 62, 554-563.

Youngstrom, E., Loeber, R., & Stouthamer-Loeber, M. (2000). Patterns and correlates of agreement between parent, teacher, and male adolescent ratings of externalizing and internalizing problems. *Journal of Consulting and Clinical Psychology*, 68, 1038-1050.

Yu, Y., & Williams, D. R. (1999). Socioeconomic status and mental health. In J. Phelan & C. Aneshensel (Eds.), *Handbook of the Sociology of Mental Health*. (pp. 151-166). Plenum.

Zelazo, P. D., & Müller, U. (2002). Executive function in typical and atypical development. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 445-469). Malden, MA: Blackwell Publishing.

Zelazo, P. D., Müller, U., Frye, D., & Stuart, M. (2003). The development of executive function in early childhood. *Monographs of the society for research in child development*, 68(3).

Zhou, Q. Hofer, C., Eisenberg, N., Reiser, M., Spinrad, T. L., & Fabes, R. A. (2007). The developmental trajectories of attention focusing, attentional and behavioral persistence, and externalizing problems during school-age years. *Developmental Psychology*, 43(2), 369-385.