The Development of and Disparities in Student Social-Emotional Competence in the Context of School-Based Social and Emotional Learning Initiatives

By

Juyeon Lee

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Social Welfare

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Valerie Shapiro, Chair
Professor Susan Stone
Professor Sophia Rabe-Hesketh

Spring 2022

Abstract

The Development of and Disparities in Student Social-Emotional Competence
in the Context of School-Based Social and Emotional Learning Initiatives

by

Juyeon Lee

Doctor of Philosophy in Social Welfare

University of California, Berkeley

Professor Valerie Shapiro, Chair

Social-emotional competence (SEC) is a set of cognitive, affective, and behavioral skills that enable an individual to engage in healthy interactions with oneself and other human beings. SEC is a critical factor for individual well-being and may also contribute to creating a more safe and caring community. SEC can develop throughout a lifetime, but it has been studied to be particularly malleable during childhood and adolescence. School-based social and emotional learning (SEL) is an interventional approach to fostering SEC in children within a school setting. In the United States and many other countries, universal SEL targeting the entire student body has become a popular practice with the goal of promoting mental, behavioral, and academic well-being of young people. Experimental, quasi-experimental, and meta-analytic research has supported the overall benefits of universal SEL on student SEC development and other positive outcomes.

However, the literature lacks sufficient evidence to answer the questions about the typical SEC developmental trajectory in the context of school-based SEL and the disparities therein across diverse student subgroups. Also, the question remains unanswered as to whether universal SEL works equally or differentially across diverse sociocultural subgroups, and the discussion of its potential impacts on subgroup disparities has been rare. Composed of three chapters each featuring a freestanding empirical study, this dissertation addresses important but under-investigated questions in the field: (1) What is a typical SEC developmental trajectory expected for school-aged children under a routine SEL practice setting? (2) To what extent do disparities exist in SEC developmental trajectories across diverse sociocultural subgroups? (3) What are the effects of universal SEL on subgroup disparities in SEC developmental trajectories?

Chapter 1 examines how student SEC develops, on average, across elementary school years under a routine SEL practice condition. Although universal progress monitoring of student SEC has increasingly been adopted as part of regular educational practices, an evidence base has

not yet been established on the typical SEC growth to expect within the school setting. In addition, longitudinal measurement invariance is a prerequisite when measuring student growth, but little evidence exists on the equivalence of SEC measurement across time, especially for the assessment tools that are currently widely used in practice. To address these gaps, this chapter examines how student SEC develops, on average, across elementary school years under a routine SEL practice condition, after first testing the longitudinal measurement invariance of a widely-used teacher-completed behavioral rating scale. The data analyzed in this chapter originate from six waves of teacher ratings of student SEC, collected from a districtwide SEL initiative for three consecutive years using the DESSA-Mini (N = 1,146; Grades K-2 at baseline). Using longitudinal confirmatory factor analysis methods, this chapter found no evidence for violation of measurement invariance across all six time points throughout the three consecutive years, suggesting that the same construct of SEC was measured across different seasons and grade levels. Then, using second-order latent growth modeling methods that did not impose any predetermined shape of growth, this chapter found that (a) student SEC increased within each year, (b) student SEC decreased over each summer by about a half of the yearly gain, and (c) the rate of yearly growth gradually decreased across years. Implications and limitations of these findings are discussed with suggestions for future research and practices.

Chapter 2 examines the extent to which disparities exist in student SEC developmental trajectories across socioculturally classified subgroups. Although student gender, race, ethnicity, and socioeconomic status (SES) have conceptually been linked to differential and unequal social-emotional developmental experiences, empirical evidence on subgroup disparities in SEC developmental trajectories has been limited and inconclusive. Also, despite the field's widespread concern about teacher bias in assessing SEC within diverse student bodies, little evidence is available on the measurement invariance of SEC progress monitoring tools across student subgroups, a prerequisite condition for examining subgroup disparities. The data analyzed in this chapter involve nine waves of teacher-rated student SEC, collected from a districtwide SEL initiative for three consecutive years (N = 5,452; Grades K-2 at baseline). Using multigroup confirmatory factor analysis methods, this chapter first tested measurement invariance of the DESSA-Mini by gender (male vs. female), race and ethnicity (Black vs. Hispanic vs. White), and SES (low-income vs. middle-to-high-income). No evidence for violation of measurement invariance across all the examined subgroups was found, suggesting that the same construct of SEC was measured across diverse student subgroups. Then, using second-order piecewise latent growth modeling methods, this chapter examined the subgroup disparities in the three-year growth trajectories of student SEC under a routine universal SEL practice condition. The results suggest that (a) male (vs. female), Black or Hispanic (vs. White), and lower-income (vs. higher-income) students started with a lower level of SEC, and (b) these gaps were either sustained across time or slightly widened. Implications and limitations of these findings are discussed with suggestions for future research and practices.

Chapter 3 examines the universality of the intervention effects of universal SEL across diverse student subgroups, and the effects of universal SEL on existing subgroup disparities. Although previous SEL research has accumulated robust evidence of the average intervention effects of universal SEL, it remains unclear whether universal SEL works equally or differentially across diverse sociocultural subgroups of students. This has implications for possible subgroup disparities in student social-emotional competence (SEC) development. The data analyzed in this chapter came from student SEC progress monitoring collected during a one-year quasi-experimental study of a universal SEL program (N = 1,592; Grades K-2). Using a multigroup structural equation model by student subgroup, this chapter first examined whether the effects of universal SEL on student SEC growth differed across diverse student subgroups classified by gender, race, ethnicity, socioeconomic status, disability status, and English learner status. The results suggest that (a) the intervention effects were slightly larger for Black students, compared to White or other racial-ethnic subgroups, and (b) the effects were not different across other examined subgroups. Using a multigroup structural equation model by intervention condition, this chapter also examined whether the patterns of any subgroup disparities in SEC growth trajectories differed by intervention condition. The results suggest that (a) in the comparison condition, the SEC disparities between Black and White students tended to widen throughout the year, whereas in the intervention condition, Black students showed a similar rate of growth as their peers, and (b) no other subgroup disparities in SEC growth trajectories showed different patterns between the two conditions. Implications and limitations of these findings are discussed with suggestions for future research and practices.

In summary, this dissertation contributes to the literature by providing evidence of (1) an average multi-year developmental trajectory of SEC among school-aged children, assessed with a measure tested to be longitudinally invariant, (2) gender, racial-ethnic, and socioeconomic disparities that exist in this SEC developmental trajectory, assessed with a measure tested to be invariant across subgroups, and (3) the potential of universal SEL for preventing some disparities from further widening, but not necessarily closing all the existing gaps.

I hope the findings of this dissertation lead to and inform larger questions in the field, such as how to make sure SEL assessment tools are accurate and equitable, how to set SEL standards and benchmarks, and how to ensure school-based SEL truly benefits all students and produces more equitable outcomes.

Table of Contents

List of Tables

Chapter 1: Measuring the Development of Social-Emotional Competence in Elementary School Years Using Behavioral Rating Scales

---

Chapter 2: Gender, Racial-Ethnic, and Socioeconomic Disparities in the Development of Social-Emotional Competence Among Elementary School Students

---

Chapter 3: Universal school-based social and emotional learning (SEL) for diverse student subgroups: Implications for enhancing equity through SEL

List of Figures

Chapter 1: Measuring the Development of Social-Emotional Competence in Elementary School Years Using Behavioral Rating Scales

---

Chapter 2: Gender, Racial-Ethnic, and Socioeconomic Disparities in the Development of Social-Emotional Competence Among Elementary School Students

---

Chapter 3: Universal school-based social and emotional learning (SEL) for diverse student subgroups: Implications for enhancing equity through SEL

Acknowledgements

I would not have been able to complete this dissertation without many people who provided invaluable support throughout my journey to a doctoral degree.

My advisor, Dr. Valerie Shapiro, I would like to thank you from the bottom of my heart. You have helped me grow so much, academically, professionally, and personally. I was able to be myself on this journey thanks to your genuine care and wholehearted support. I am so fortunate to have you as a mentor in my life. I look forward to sharing many more rainbows with you in the future.

I am deeply grateful for my committee members who provided valuable feedback on my work and helped me make it through the rocky times. Dr. Susan Stone, your insights and suggestions always broadened my view and made me ask hard questions. Dr. Sophia Rabe-Hesketh, your feedback on my analysis and interpretation was crucial to improve the quality of my work. Also, to both of you, I cannot say thank you enough for helping me navigate through the many uncertainties in the last few years and reach this meaningful milestone.

I wish to thank other UC Berkeley faculty members who willingly provided constructive feedback on my work and shared kind words of encouragement. Thank you to Drs. Jill Berrick, Tina Sacks, Yu-Ling Chang, Erin Kerrison, Julian Chow, Kurt Organista, Paul Sterzing, and Adrian Aguilera.

I also wish to thank my mentors at Seoul National University. Dr. Joan Yoo, thank you for believing in me even when I couldn't. You will forever be my mentor. Dr. Bong Joo Lee, I am deeply grateful for your unwavering support and encouragement. Dr. Hye-Lan Kim, the strength-based perspective I learned from you laid the foundation for my doctoral study. Dr. Sang Kyoung Kahng, this dissertation would not have been possible without your teaching of structural equation modeling. Dr. Jung-Hwa Ha, thank you for your feedback on framing my work for the broader audience.

I would like to express my sincere gratitude to Paul LeBuffe and Jennifer Robitaille at Aperture Education who provided an essential help at various stages of this dissertation project. Thank you for sharing the data and providing valuable feedback on my drafts. I would also like to acknowledge the students, teachers, administrators, staff, and funders who were part of the original projects.

I am deeply appreciative of current and previous colleagues at the Center for Prevention Research in Social Welfare. Dr. Bo-Kyung Elizabeth Kim, thank you for answering all the questions I had on the data and providing constructive feedback on my drafts. Thank you to Kelly Ziemer, Nehal Eldeeb, Dr. Emily Campbell, Dr. Stephanie Machado, and Dr. Sophia Hwang, who have been wonderful thought partners and cheerleaders to each other along this journey. I could not have made it without all your feedback and encouragement.

I cannot talk about my doctoral experiences without mentioning my dear unicorn ladies: Katie Savin, Jaclyn Chambers, Cristina Gomez-Vidal, Mayra Cazares-Minero, and Kelly Ziemer. Each of you were a tremendous source of inspiration, support, and joy for me. I am so lucky to be part of this legendary cohort.

I also wish to thank my Korean friends who provided incredible support and encouragement, especially during the difficult times. Thank you to Woojin Jung, Eunkyung Chung, MinJee Kye, Mina Lee, Sol Baik, In Young Park, Hyojin Cho, Eunseok Jeong, Joonmo Kang, and Chung Hyeon Jeong.

To my parents, thank you for your unending love and support. Juyi and Juseok, thank you for being there for me. Jueun, you are forever my muse. I would also like to thank my parents in law for being so supportive throughout this journey.

Lastly, thank you to Dr. Changyong Choi – my best mentor, colleague, friend, and love – especially for your patience and sacrifice for me and Jin. And to Jin, thank you for being in this world and in my life. You are now the strongest motivation for me to contribute to creating a more safe, caring, and just society.

Introduction

Social-emotional competence (SEC) is defined as the knowledge, attitudes, and skills necessary to know and manage the self, understand and relate with others, and make responsible personal and social decisions (Elias et al., 1997; Weissberg et al., 2015). SEC is a multidimensional construct, involving intrapersonal and interpersonal competencies, such as self-awareness, social-awareness, self-management, relationship skills, and responsible decision-making (Collaborative for Academic, Social, and Emotional Learning, 2022b). It is suggested that achievement of these competencies requires functional integration across cognitive, affective, and behavioral systems (Domitrovich et al., 2017).

SEC is theorized to be crucial to healthy development and well-being for children and youth for several reasons. First, SEC has been demonstrated to act as a protective factor that can moderate the negative influences of various individual and environmental risk factors on developmental outcomes (Domitrovich et al., 2017). Second, SEC has been shown to be a reliable predictor of educational success, including school readiness in early childhood (Denham, 2006) and academic achievement in middle childhood and adolescence (Zins et al., 2004). Third, SEC is described as a promotive factor of individual well-being: from a hedonic perspective, SEC facilitates alleviation of psychological distress and occurrence of positive affect; from a eudaimonic perspective, it sets conditions for fulfillment of basic psychological needs for competence, relatedness, and autonomy (Ryan et al., 2008). Furthermore, SEC has a potential to contribute to societal and collective well-being, because it promotes individual children's prosocial and socially responsible behaviors, which in turn may facilitate positive peer interactions and create a caring and respectful community that provides a context for positive development for all youth (Elias et al., 1997; Ryan et al., 2008).

The development of SEC occurs at various developmental stages and through various socialization contexts. In infancy and toddlerhood, children develop a sense of self and others, emotional and behavioral control, concerns for others and prosocial behaviors, as well as an understanding of social conventions and norms (Shulman, 2016; Wittmer et al., 1996). In this early stage of life, the family environment plays a critical role in children's social and emotional development. As children age, schools become another important context for the development of SEC. School entry provides a new set of opportunities and challenges for children's social and emotional development, as they enter into a distinctive social world from earlier socialization experiences and spend a significant amount of their time in schools (W. A. Collins, 1984). The literature on social and emotional development and learning suggests that SEC is particularly malleable throughout the school years and can be promoted through effective school-based interventions (Borst & Srinivasan, 2022; Domitrovich et al., 2017).

Social and emotional learning (SEL) refers to the process of acquiring and developing SEC (Elias et al., 1997; Weissberg et al., 2015). Although a variety of approaches to SEL exist in practice, SEL has typically been featured as a classroom-based intervention intended to help students process, develop, integrate, and apply SEC in developmentally and contextually appropriate ways, as well as to foster positive learning environments that are caring, supportive, and engaging (Durlak & Weissberg, 2011; Osher et al., 2016).

Prevention scientists have suggested that school-based SEL is a promising intervention for preventing a wide range of youth behavioral health problems, including violence, delinquency, substance use, and school failure (Domitrovich et al., 2017; Elias et al., 2015; D. J. Jones et al., 2015). Universal preventive interventions, of which the target population is the entire student body, constitute the most typical approach to cultivating SEL in schools (Farrell & Vulin-Reynolds, 2007; Muñoz et al., 1996). School-based, universal SEL that seeks to reduce the risks of various problems by leveraging the overall level of student SEC has been suggested to yield greater public health benefits than other selective and indicated prevention strategies (i.e., those that target a subgroup of students based on the level of risk for a problem or early signs of atypical development), without stigmatizing high-risk students (Greenberg et al., 2017).

The interest in SEL, however, is not limited to the field of prevention. In the positive youth development tradition, school-based SEL programs also seek to create positive developmental pathways, in addition to prevention of problems (Catalano et al., 2002; Domitrovich et al., 2017). For educators, SEL is aligned with whole-child education approaches (Lewallen et al., 2015). For economists and private-sector employers, SEL is also seen as a tool that can be used to prepare students with "21st century skills" or "life skills" that are transferable to work and life after schooling (Kautz et al., 2014; National Research Council, 2012; World Health Organization, 1994). From these various perspectives, school-based, universal SEL is advocated as a promising approach to addressing the grand challenge for social work to ensure healthy development for all youth (Hawkins et al., 2015).

Over the past two decades, school-based SEL has been most widely studied in the United States and other Western countries, primarily in elementary and middle schools. Hundreds of evaluation studies have reported the effects of SEL programs on student SEC as well as other behavioral and academic outcomes (Weissberg et al., 2015). Several meta-analyses have summarized the overall effects of school-based universal SEL programming, tested through experimental or quasi-experimental designs (Corcoran et al., 2018; Durlak et al., 2011; Sklad et al., 2012; Taylor et al., 2017). These studies found that SEL had significant positive effects, on average, on promoting SEC and other positive student outcomes. These studies have led to the declaration of an "evidence base" for SEL as a general practice.

However, the SEL literature lacks sufficient evidence to answer the questions about the typical SEC developmental trajectory in the context of school-based SEL and the disparities therein across diverse student subgroups. Also, the question remains unanswered as to whether universal SEL works equally or differentially across diverse sociocultural subgroups and its implications on subgroup disparities and equity in SEL outcomes. Composed of three freestanding chapters, this dissertation addresses important but under-investigated questions in the field: (1) What is a typical SEC developmental trajectory expected for school-aged children under a routine SEL practice setting? (2) To what extent do disparities exist in SEC developmental trajectories across diverse sociocultural subgroups? (3) What are the effects of universal SEL on subgroup disparities in SEC developmental trajectories?

Chapter 1

Measuring the Development of Social-Emotional Competence in Elementary School Years Using Behavioral Rating Scales

---

Summary

Universal progress monitoring of student social-emotional competence (SEC) has increasingly been adopted as part of regular educational practices in the context of schoolwide social and emotional learning (SEL). However, an evidence base has not yet been established on the typical SEC growth to expect within the school setting. In addition, although longitudinal measurement invariance is a prerequisite when measuring student growth, little evidence exists on the equivalence of SEC measurement across time, especially for the assessment tools that are currently widely used in practice. To address these gaps, this chapter presents a study that examines how student SEC develops, on average, across elementary school years under a routine SEL practice condition, after first testing the longitudinal measurement invariance of a widely-used teacher-completed behavioral rating scale. The data analyzed in this study originate from six waves of teacher ratings of student SEC, collected from a districtwide SEL initiative for three consecutive years using the DESSA-Mini (N = 1,146; Grades K-2 at baseline). Using longitudinal confirmatory factor analysis methods, this study found no evidence for violation of measurement invariance across all six time points throughout the three consecutive years, suggesting that the same construct of SEC was measured across different seasons and grade levels. Then, using second-order latent growth modeling methods that did not impose any predetermined shape of growth, this study found that (a) student SEC increased within each year, (b) student SEC decreased over each summer by about a half of the yearly gain, and (c) the rate of yearly growth gradually decreased across years. Implications and limitations of these findings are discussed with suggestions for future research and practices.

Background

Social and emotional learning (SEL) refers to the process of acquiring and developing social-emotional competence (SEC), which has been defined as the knowledge, attitudes, and skills necessary to know and manage the self, understand and relate to others, and make responsible personal and social decisions (Elias et al., 1997; Weissberg et al., 2015). SEC is a multidimensional construct that involves distinct but highly interrelated cognitive, emotional, and behavioral competencies that facilitate effective functioning in intrapersonal and interpersonal interactions (Buckley & Saarni, 2009). In the United States, school-based SEL has become a popular approach to Tier 1 interventions (in education terminology) or universal prevention (in public health terminology), based on SEL's conceptual and empirical linkages with the prevention of behavioral problems (Domitrovich et al., 2017; Elias et al., 2015), whole-child education (Lewallen et al., 2015), positive youth development (Catalano et al., 2002), and life skills development (Kautz et al., 2014; National Research Council, 2012). With research evidence supporting the overall benefits of school-based universal SEL (e.g., Corcoran et al., 2018; Durlak et al., 2011; Sklad et al., 2012; Taylor et al., 2017), recent policy and funding contexts have created, more than ever, favorable conditions for districtwide and statewide adoption of SEL (Gayl, 2017; Krachman & LaRocca, 2017).

Accordingly, there has been a growing need for universal SEL assessments in practice, especially for formative assessments that monitor student progress in SEC (Hamilton & Doss, 2020). Universal SEC progress monitoring data can not only formatively guide classroom SEL practices, but also lay the foundation for system-level accountability and decision-making to guide planning, implementation, and evaluation of schoolwide SEL initiatives (Grant et al., 2017; LeBuffe et al., 2018; McKown, 2019a). These data may also be used to identify individual students who would benefit from expanded interventions in a multi-tiered system of support, for example, based on their lower baseline level of SEC and/or slower rate of growth under a universal SEL intervention condition (Ardoin et al., 2005; LeBuffe et al., 2018; Verlenden et al., 2021).

To make the best use of SEC progress monitoring data, educators might benefit from having a reference for "typical growth" to inform how much growth to expect in their students over time under a routine SEL practice condition. However, the current literature has not yet sufficiently accumulated such evidence to guide practice decisions with diverse samples of school-aged children. The emergence and development of self-regulatory and prosocial behaviors in early childhood has long been an important topic in developmental science. The literature suggests that, overall, toddlers and preschoolers tend to show more socially and emotionally competent behaviors as they grow older (Hay et al., 2021; National Scientific Council on the Developing Child, 2007; Rose-Krasnor & Denham, 2009; Yates et al., 2008), although the observed extent of changes has varied across studies (Eisenberg et al., 2006; Spinrad & Eisenberg, 2009; Zahn-Waxler & Smith, 1992). Less research has been conducted on social-emotional developmental trajectories during the elementary school years and beyond, and previous studies tended to focus more on measuring specific sub-constructs of SEC (e.g., emotion regulation, prosocial behavior) rather than the overall level of SEC (Buckley & Saarni, 2009).

Although there have been some noteworthy studies of prosocial developmental trajectories in middle childhood (e.g., Aber et al., 2003; Côté et al., 2002; Kokko et al., 2006; Sørlie et al., 2020), findings seem to be inconclusive as to the direction and shape of changes. For example, some studies have found an overall declining pattern of teacher-rated prosocial behavior from age 6 to 12 (Côté et al., 2002; Kokko et al., 2006), while one found that teacher-rated social skills showed a linearly increasing pattern, on average, from grade 4 to 7 (Sørlie et al., 2020). In another study, teacher-rated prosocial behavior showed a slightly decreasing pattern from age 6 to age 7.5, then an accelerated increase up to age 12.5 (Aber et al., 2003). These inconsistencies are likely due to the differences in sample characteristics and measures in use, calling for more studies with diverse student populations and instruments that are tested to be valid for measuring progress over time in school-aged children.

The widescale adoption of universal assessments of student SEC in educational practice may reciprocally produce useful data that can contribute to our knowledge of how children's SEC develops within and across school years, under routine SEL practice conditions, across various contexts. A more robust understanding of average SEC developmental trajectories can also contribute to our predictions of the variations in these trajectories at the individual student level and at the classroom, school, district, or state level, as well as how these trajectories relate to various student outcomes in a longer term. Such knowledge, in turn, can inform practitioners of the benchmarks they would expect their students to reach in response to an intervention in practice, which will guide evidence-based screening, instructional support, and progress monitoring.

Teacher-completed Behavioral Rating Scales in SEC Assessments

For the aforementioned reasons, there has been an increasing demand for assessment tools that can be used universally and repeatedly to measure student SEC in routine school settings. Teacher-completed behavioral rating scales and questionnaires are one of the most commonly used tools in SEC assessments because of their relative utility compared to other measurement methods (Melnick et al., 2017). Teacher-completed behavior rating scales typically consist of a series of items asking teachers to rate the perceived frequency of a student's behavior indicative of SEC during a given period (e.g., the past four weeks) on a Likert scale (e.g., *never* to *very frequently*). Some questionnaires also include similar Likert-type questions asking about teachers' perception of a student's skills and attitudes demonstrating SEC. Despite concerns regarding the indirect and subjective nature of observations, teachers are usually thought to be good informants of student behavior in the natural school context, with a strong reference group to use for rating individual students' behavior (Assessment Work Group, 2019; Duckworth & Yeager, 2015). Compared to some resource-intensive methods such as performance tasks, direct observations, or structured interviews, teacher-completed behavioral rating scales tend to be more practical (e.g., easy to administer, time-efficient, and cost-efficient), particularly when ratings scales are brief (Anthony et al., 2020; Assessment Work Group, 2019; Duckworth & Yeager, 2015). Compared to student self-reports, teacher-completed behavioral rating scales or questionnaires may be advantageous with younger students, avoiding concerns about literacy or introspective ability, and may reduce social desirability bias as children become increasingly aware of social norms and expectations as they grow older (Assessment Work Group, 2019; Duckworth & Yeager, 2015; Zahn-Waxler & Smith, 1992). For these reasons, teachers often provide ratings of student behavior, especially in elementary school settings where the primary classroom teacher spends a substantial amount of time with the same students (LeBuffe et al., 2018).

A teacher-completed behavioral rating scale or questionnaire needs to be brief and psychometrically sound in order to be used as a universal SEC assessment tool in routine school settings. There is no consensus on the reasonable number of items for a brief tool, but it has been suggested that teachers should be able to rate an entire classroom of students in a single prep period, implying an administration time of 1-2 minutes per student for a typical classroom size (LeBuffe et al., 2018). One study found that student behavioral progress could be effectively monitored with as few as 12 items (Gresham et al., 2010). SEL assessment tools should also demonstrate evidence of reliability and validity for their intended use (American Educational Research Association et al., 2014). Although the required levels of evidence for a SEC progress monitoring tool might be relatively lower than higher-stakes assessments (Assessment Work Group, 2019), sufficient evidence of reliability and validity for its intended uses should be available.

Currently, few assessment tools of student SEC have been judged to be both sufficiently brief and psychometrically defensible. According to the SEL Assessment Guide (Assessment Work Group, 2021), one of the most up-to-date catalogs of SEC assessments created by a multidisciplinary team of researchers and practitioners, eight teacher-reported measures of student SEC are actively used in educational practice (i.e., not intended for clinical use or risk screening purposes). Among these eight tools, five are behavioral rating scales or questionnaires that can be used across different SEL programmatic approaches and can produce rating scores at the individual level (i.e., excluding a rubric-type measure asking teachers to assess a global level of student SEC rather than a set of specific behaviors or skills demonstrating SEC, another that was specifically designed to be used with a particular SEL program, and another only reporting aggregated scores at the group level). Among these five tools, three involve more than 50 items; two are brief: (1) an 8-item behavioral rating scale, the Devereux Student Strengths Assessment Mini (DESSA-Mini; Naglieri et al., 2011/2014) and (2) a 10-item survey questionnaire, the Panorama Teacher Perceptions of Students' SEL (Panorama Education, 2017). However, at the time of this writing, evidence of psychometric properties is only available for the DESSA-Mini (see LeBuffe et al., 2018). No peer-reviewed publications or other publicly available reports of the Panorama teacher survey's psychometric qualities could be located.

The DESSA-Mini is, in fact, designed to be a practical and psychometrically sound measure of SEC that can be used for universal screening and progress monitoring in the context of a large-scale implementation of SEL (LeBuffe et al., 2018). As a short version of the full 72-item DESSA (LeBuffe et al., 2009/2014), the DESSA-Mini is a norm-referenced behavioral rating scale, developed with a nationally representative standardization sample of K-8 students. It has demonstrated strong psychometric qualities including reliability (e.g., internal reliability, test-retest reliability, inter-rater reliability, alternate form reliability) and validity (e.g., concurrent and predictive criterion validity), as summarized in LeBuffe et al. (2018). Although there seems to be increasing research interest in developing additional brief and psychometrically sound measures of SEC (e.g., Anthony et al., 2020), the DESSA-Mini is one of the very few such tools that are currently available and is already in widespread use, estimated to assess approximately half a million children each year at the time of writing.

Longitudinal Measurement Invariance in SEC Assessments

One important but underinvestigated psychometric characteristics of SEC assessment tools is longitudinal measurement invariance, or equivalence of measurement across different time points. An assessment tool intended to be used for SEC progress monitoring should be able to accurately measure the change in student SEC (Assessment Work Group, 2019).

Metrics of change sensitivity often include increased scores as a function of student chronological age and/or interventions in place, but this approach relies on the assumption that the same construct is measured at each occasion. This assumption of measurement invariance across time, however, may be violated in reality. For example, even when the same teacher rates the same student's SEC over time using the same behavioral rating scale, if the teacher's conceptualization of SEC, or the ways in which they rate the same behavior, change over time, it would be incorrect to interpret the observed changes in rating scores as actual changes in student SEC. Examination of multi-year growth using a behavioral rating scale can be even more challenging as students are usually rated by different teachers across years.

For a multi-item instrument, longitudinal measurement invariance can be empirically tested by examining whether the structural relation between observed indicators and the latent factor is constant across time. In the structural equation modeling framework, this is done by evaluating four levels of factorial invariance: configural invariance (the same set of indicators for the latent factor are specified at each occasion), weak factorial invariance (the factor loading of each indicator is invariant across time), strong factorial invariance (the factor loading and intercept of each indicator are invariant across time), and strict factorial invariance (the factor loading, intercept, and unique variance for each indicator are invariant across time; Ferrer et al., 2008; Widaman et al., 2010). Developmental scientists have emphasized the importance of testing longitudinal measurement invariance in any studies of human behavior change or growth (Millsap, 2010; Widaman et al., 2010), and have demonstrated that the lack of strong or strict factorial invariance can lead to erroneous conclusions about the patterns of growth in latent growth modeling (Ferrer et al., 2008).

Longitudinal confirmatory factor analysis (CFA) has been suggested as a statistical method of testing factorial invariance across time, which appropriately takes into account the longitudinal nature of the data (versus a multigroup CFA approach) and can easily be extended into second-order latent growth modeling that accommodates individual growth curves while accounting for measurement errors (Vandenberg & Lance, 2000; Widaman et al., 2010). Despite the availability of the modeling techniques that can be used to accurately measure student SEC growth, investigation of longitudinal measurement invariance appears to have been rare in the field of SEC assessment. A handful of recent studies have tested and confirmed factorial invariance of a newly developed measure of student SEC across two or three time points (Anderson-Butcher et al., 2016; McKown, 2019b; Peck et al., 2018; Ross & Tolan, 2018). But only a couple of these studies employed the longitudinal CFA approach suggested by the literature on measurement invariance, confirming factorial invariance of a self-report measure of SEC (Anderson-Butcher et al., 2016; Ross & Tolan, 2018). To my knowledge, no evidence exists on the longitudinal measurement invariance of a teacher-rated behavioral rating scale used for SEC assessment, particularly the ones that are already widely used in universal progress monitoring, such as the DESSA-Mini.

Research Objectives

The objective of this study is twofold: (1) to test longitudinal measurement invariance of a teacher-completed behavioral rating scale of student SEC, and (2) to examine elementary students' average SEC growth trajectory in the context of school-based SEL. Specifically, this study examines the four levels of factorial invariance of a widely used SEC rating scale (DESSA-Mini) across multiple occasions within and across school years, and pending at least strong factorial invariance over time, estimates the average growth trajectory of student SEC throughout three elementary school years in the context of a district-wide SEL initiative.

As reviewed earlier, the literature lacks consistent evidence to inform a hypothesis about the direction and shape of age-related growth in middle childhood. However, given the study context where a district-wide SEL initiative was in place, an increasing rather than decreasing pattern of SEC is hypothesized. But when it comes to the shape of a growth curve, no specific hypothesis was made given the lack of prior research.

Methods

Practice Context

Data used in this study come from a district-wide SEL initiative that implemented the Promoting Alternative Thinking Strategies (PATHS) curriculum (Kusché & Greenberg, 1994) in a large urban district for three academic years from 2011-2012 to 2013-2014. This district serves a racially and ethnically diverse and low-income student body in Pennsylvania, characterized by a high mobility rate. During the study years, the district's average elementary school enrollment was 8823 students, 64.5% of whom were identified as Hispanic, 14.8% non-Hispanic Black, and 13.2% non-Hispanic White, with approximately 86% receiving free or reduced-price meals, and 42% transferring in and out of schools (Pennsylvania Department of Education, 2021). The PATHS curriculum is a classroom-based universal SEL program found to be effective in promoting social-emotional, behavioral, and academic outcomes among school-aged children in several randomized trials (e.g., Conduct Problems Prevention Research Group, 1999; Fishbein et al., 2016; Greenberg et al., 1995; Schonfeld et al., 2015). The PATHS curriculum was implemented with K-2 students across 15 elementary schools in Year 1 (2011-2012), and with all K-5 students across these 15 elementary schools in Year 2 (2012-2013) and Year 3 (2013-2014) with the exception of three schools that elected to discontinue implementing PATHS starting Year 2. A full-day initial training and a half-day booster training were provided to K through 2nd grade teachers in Year 1, and to 3rd through 5th grade teachers in Year 2. Teachers were also provided with ongoing support by PATHS coaches who conducted regular classroom visits (monthly in Year 1, bi-monthly in Year 2, and as requested in Year 3). Despite variability across teachers and schools, and increasing sparsity of third-party observations across Year 1, the PATHS program was delivered with a high degree of fidelity and commitment in Year 1, as indicated by teachers' self-reports as well as coaches' ratings (Fleming, 2014; Shapiro et al., 2018). No published data is available on the extent to which the program was implemented in Year 2 and Year 3. Study protocols were approved by the Institute of Clinical Training and Research at the Devereux Foundation. De-identified data (no longer regarded as human subjects data) was subsequently shared with the first author, whose analytic activities were acknowledged by the Committee on Protection of Human Subjects at the University of California, Berkeley.

Measurement

The DESSA is a standardized, norm-referenced, behavioral rating scale that assesses K-8 students' SEC (Naglieri et al., 2013), theorized to consist of eight domains (self-awareness, social awareness, self-management, goal-directed behavior, relationship skills, personal responsibility, decision making, and optimistic thinking), which aligns well with the leading SEL framework in the United States developed by the Collaborative for Academic, Social, and Emotional Learning (2020). The full DESSA comprises 72 items asking teachers to rate the frequency of a student's behavior observed during the past four weeks on a 5-point Likert scale (0=Never, 1=Rarely, 2=Occasionally, 3=Frequently, 4=Very frequently). All of the items are strengths-based, indicating positive rather than maladaptive behaviors (Simmons et al., 2016). Example items include "(how often did the child) keep trying when unsuccessful," "respect another person's opinion," "do something nice for somebody," and "show good judgment." In practice, item raw scores are transformed into T-Scores based on a nationally representative standardization sample (M=50, SD=10) to be used to assess individual students' relative social-emotional strengths and needs: T-Scores of 60 and above indicate that the student has *Strengths*, T-Scores between 41 and 59 represent *Typical* SEC, and T-Scores of 40 and below indicate a *Need for Instruction*.

The DESSA-Mini is a brief, 8-item version of the full DESSA that measures the overall level of SEC. There are four DESSA-Mini forms, each containing a set of 8 non-overlapping items from the full set of 72 items on the DESSA. The DESSA-Mini items that were selected from the full set of DESSA items were the items that most strongly correlated with the DESSA composite T-Score, yielded similar scores across forms, and maintained some breadth of constructs. It is recommended that these forms be used in rotation when monitoring student progress in order to limit the practice effects that might occur when the small number of same items are used repeatedly over time (LeBuffe et al., 2018). In many practice settings, including the district-wide initiative described in this study, the DESSA-Mini Forms 1 to 3 have been used to measure the growth of SEC throughout the year in order (Naglieri et al., 2011).

Throughout the three school years of this SEL initiative, teachers assessed all of their classroom students' SEC using three DESSA-Mini Forms each year (Form 1 in Fall, Form 2 in Winter, and Form 3 in Spring). In addition, teachers were asked to complete the full DESSA in Fall (around October) and Spring (around June) each year on a randomly selected 5-8 students in their class. In the current study, the 24 items comprising the DESSA-Mini Forms 1 to 3 were extracted from the full DESSA, as consistent with the method articulated by Shapiro, Kim, et al. (2017), this time to test longitudinal factorial invariance of each DESSA-Mini form across six occasions (Fall and Spring each year for three consecutive years) and to estimate the average growth trajectory of SEC measured by each DESSA-Mini form throughout the study years.

The DESSA-Mini has demonstrated strong psychometric qualities including reliability (e.g., internal reliability, test-retest reliability, inter-rater reliability, alternate form reliability) and validity (e.g., concurrent and predictive criterion validity), as evidenced in several studies (Lee et al., 2018; Shapiro, Accomazzo, et al., 2017; Shapiro, Kim, et al., 2017). In the original standardization study underlying the development of the DESSA-Mini (Naglieri et al., 2011/2014), each of the three DESSA-Mini forms showed evidence of excellent reliability: the internal reliability coefficient estimates (Cronbach's alpha) ranged from .919 to .924, the test-retest reliability coefficient estimates ranged from .88 to .94, the inter-rater reliability coefficient estimates ranged from .77 to 81, and the alternate form reliability coefficient estimates ranged from .92 to .93 across all pairs of three forms. In the current study, the internal reliability coefficient estimates were higher (.94 to .98) for all three forms across all six occasions, as was the case in the DESSA-Mini validity study that analyzed extracted items from the full DESSA (Shapiro, Kim, et al., 2017). The alternate form reliability coefficient estimates were also higher (.95 to .98) than the original standardization study across all pairs of composite scores at all six occasions. (See Table 1.1). In the study sample, over 90% of the students were rated by the same teachers within a single year, and over 96% by different teachers across years.

Table 1.1. Internal and Alternate Form Reliability Estimates of the Items Comprising the DESSA-Mini (Forms 1-3; 8 items each)

Time	Cronbach's alpha			Alternate Form Reliability		
	Form 1	Form 2	Form 3	Form 1&2	Form 1&3	Form 2&3
Year 1 Fall	.94	.94	.95	.95	.95	.95
Year 1 Spring	.94	.94	.95	.96	.95	.95
Year 2 Fall	.96	.96	.96	.97	.96	.96
Year 2 Spring	.97	.97	.97	.98	.97	.97
Year 3 Fall	.97	.97	.97	.97	.97	.96
Year 3 Spring	.97	.97	.98	.98	.97	.97

Study Sample

This study focuses on students who were in grades K-2 in Year 1 (grades 2-4 in Year 3) to examine the growth trajectory of SEC over three consecutive elementary school years under the routine implementation of an SEL intervention. Based on the finding suggesting the absence of age trends in the DESSA-Mini ratings across Grades K-8 (Naglieri et al., 2011/2014), students of different grade levels were considered as consisting of a single sample representing children in the same developmental stage (i.e., middle childhood). The sample of the current study consists of 1,146 students who were randomly selected to be rated on the full DESSA. In Year 1, five to eight students per class were randomly selected to be rated on the full DESSA. In the following years, an attempt was made to assess the same students using the full DESSA. When a given teacher had fewer than five students to rate in their class, additional students were randomly selected so that all teachers were asked to rate five students. In Year 3, five schools opted out of rating randomly selected students on the full DESSA, but instead started using the full DESSA to assess students identified as having higher social-emotional needs per their DESSA-Mini ratings (consistent with recommended use of the DESSA in practice). Students rated in these five schools in Year 3 were excluded from the study sample.

In the study sample, grade and gender were quite evenly distributed. Students were 35.69% Kindergartners, 28.88% first graders, and 35.43% second graders at baseline as well as 47.38% female and 52.62% male. The sample was representative of the district's student population in terms of the distribution of race and ethnicity as well as SES (indicated by free or reduced-price meals eligibility): 61.87% identified as Hispanic, 15.79% Black, 15.97% White, 4.45% Multi-Race, and 1.92% Others (including American Indian or Alaska Native, Asian, and Hawaiian or Pacific Islander), and 86.91% were low-income (eligible for free or reduced-price meals). Most students (75.22%) attended PATHS-implementing schools throughout the entire study period, while about a quarter of the students (24.78%) attended one of the schools that discontinued the implementation of PATHS in Year 2 and Year 3. (See Table 1.2).

Table 1.2. Sample Characteristics

Variable	Category	Frequency	Percent
Grade at Baseline	K	409	35.69
	1	331	28.88
	2	406	35.43
Gender	Female	543	47.38
	Male	603	52.62
Race and Ethnicity	Black (non-Hispanic)	181	15.79
	Hispanic (of any race)	709	61.87
	White (non-Hispanic)	183	15.97
	Others	22	1.92
	Multi-race	51	4.45
Socioeconomic Status	Low-income	996	86.91
	Middle-to-high income	150	13.09
Exposure to the PATHS implementation	Continued for three years	862	75.22
	Discontinued in Years 2-3	284	24.78
Total		1,146	100.00

Analysis Plan

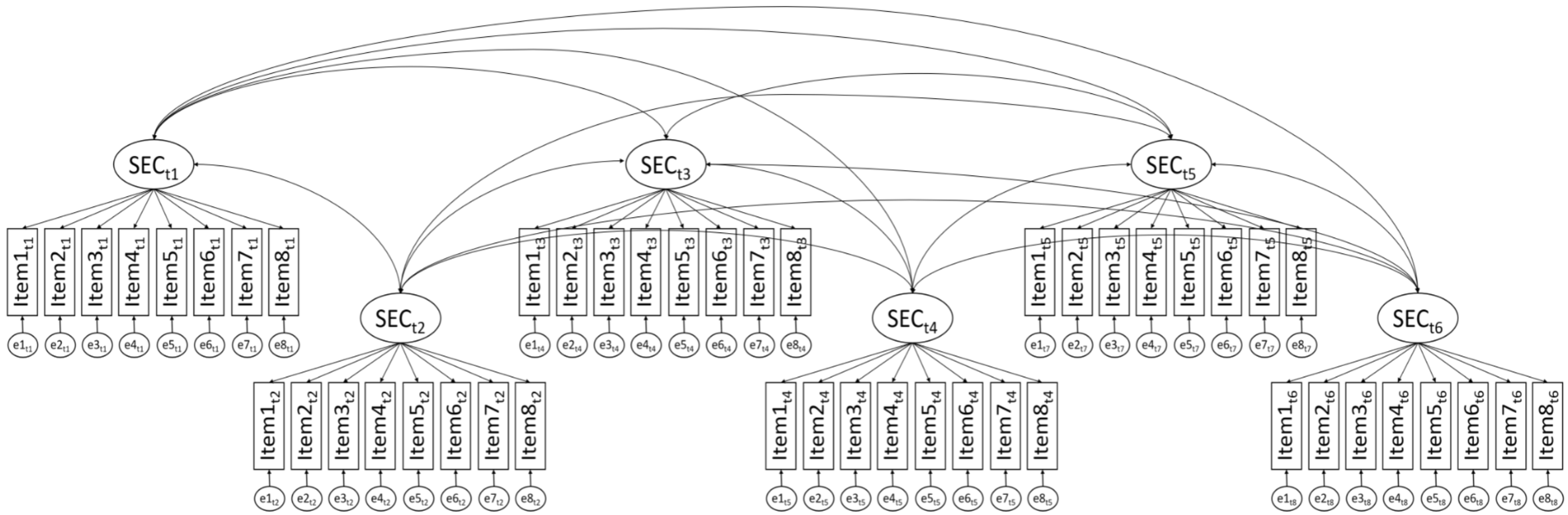
Longitudinal Factorial Invariance Testing.

For each set of eight items that comprise the DESSA-Mini Forms 1 to 3, four levels of factorial invariance (i.e., configural invariance, weak factorial invariance, strong factorial invariance, and strict factorial invariance) were tested using longitudinal CFA, following the steps articulated in Widaman et al. (2010). The longitudinal confirmatory factor model examined in this study is presented in Figure 1.1. This model consists of eight indicator variables (Item1 to Item8) measuring a single latent variable (SEC) at each of six times of measurement (t1 to t6). The SEC latent variables were allowed to covary across time as indicated by double-headed, curved arrows in Figure 1.1. The residuals for the same indicator (e1 to e8) were also allowed to covary across times of measurement although not shown in Figure 1.1 for graphical simplicity. The SEC latent mean was allowed to be different at each time point.

The configural invariance model had minimal identification constraints placed on the model presented in Figure 1.1, specifying the identical configuration of the factor-indicator relations across time: the mean and variance of the latent variable at Time 1 (SEC_{t1}) was fixed to 0 and 1, respectively, and the factor loading and intercept of the first item ($Item1_{t1}$ to $Item1_{t6}$) were estimated but constrained to be equal across time. The weak factorial invariance model added across-time equality constraints on the factor loadings for all the same indicator variables (i.e., metric invariance) to the configural invariance model. The strong factorial invariance model added across-time equality constraints on the intercepts of the same indicator variables (i.e., scalar invariance) to the weak factorial invariance model. Finally, the strict factorial invariance model added across-time equality constraints on the unique variances (i.e., residual invariance) for each of the indicator variables to the strong factorial invariance model.

The four invariance models represent a series of nested models with increasing restrictions or constraints. Several model fit indices were compared to examine the degree to which these constraints held with data. The chi-square difference test based on the likelihood ratio test statistics was used to test the statistical significance of the model fit differences. Since the chi-square statistic is very sensitive to sample size, however, alternative fit indices including the comparative fit index (CFI), the Tucker-Lewis index (TLI), and the root mean squared error of approximation (RMSEA) were also used. Given the large sample size (N = 1,146) of the current study, when the chi-square difference test suggested a significantly worse fit for a more restrictive model, but no appreciable change was found in alternative fit indices (e.g., ΔCFI smaller than or equal to $|0.01|$; Cheung & Rensvold, 2002), I opted to accept the more restricted model, as suggested by Widaman et al. (2010).

Figure 1.1. Longitudinal Confirmatory Factor Model of Student Social-Emotional Competence (SEC) Measured by Eight DESSA-Mini Items Across Six Times of Measurement



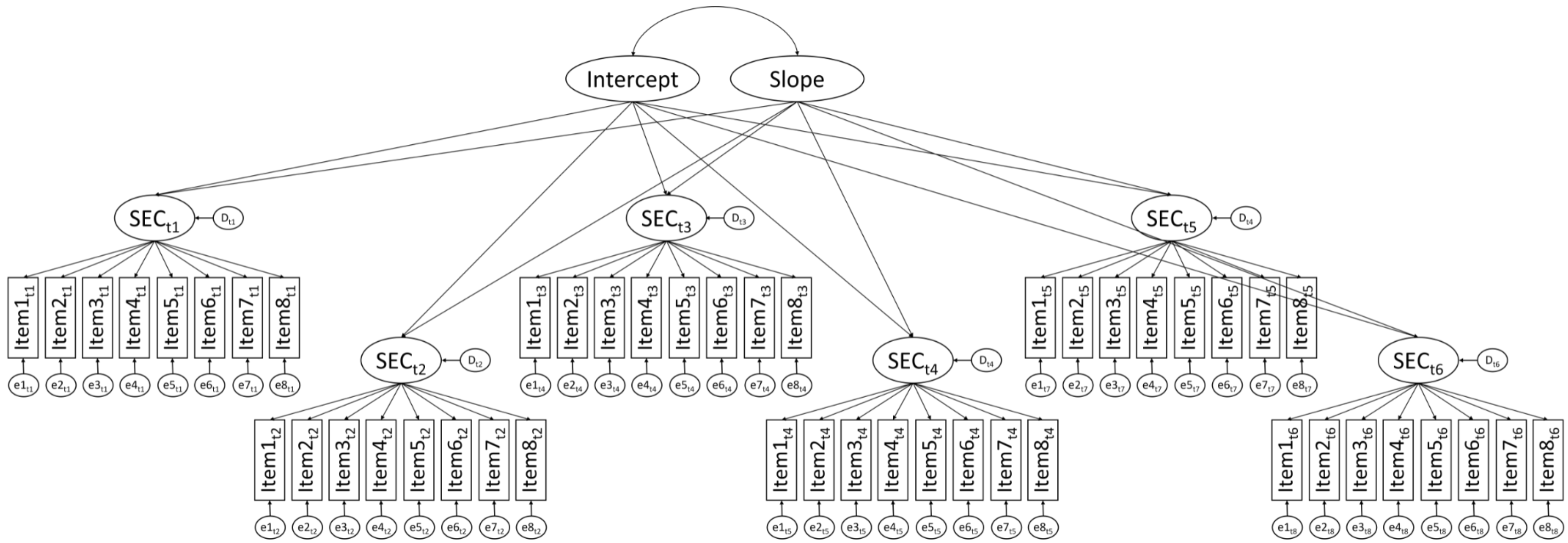
Growth Trajectory Modeling.

After confirming the comparability of latent means across time with at least strong factorial invariance, the average growth trajectory of student SEC was estimated using a second-order latent growth model that directly incorporates the measurement model into the latent growth model (T. E. Duncan & Duncan, 2004; Meredith & Tisak, 1990). The second-order latent growth model examined in this study is presented in Figure 1.2. The first-order component of this model is identical to the measurement model shown in Figure 1.1, assuming strong factorial invariance across time and the residuals for the same indicator allowed to covary across time. The second-order component of the growth model consists of two growth factors, intercept (baseline level) and slope (growth rate). Instead of assuming a linear or other predetermined shape of the growth curve, only the first two coefficients (or loadings) of the slope factor were fixed at 0 and 1, respectively, for the purpose of model identification, and the remaining four coefficients were freely estimated. The coefficients of the intercept factor were all fixed at 1, and the two growth factors were allowed to covary.

To identify the model and scale the latent variables, the latent standardization identification method was used so that the SEC latent variable at time 1 (SEC_{t1}) had the mean of 0 and the standard deviation of 1. Unlike other identification and scaling methods that can yield different parameter estimates depending on the specific method applied (e.g., depending on the choice of item in the marker variable approach), the latent standardization identification method yields unique growth parameter estimates and provides a more meaningful interpretation of the mean of the slope factor (Yang et al., 2020). With this approach, the mean of the intercept factor is fixed at 0, and the mean of the slope factor indicates the mean change in the latent SEC construct between Time 1 and Time 2 relative to the standard deviation of the latent SEC construct at Time 1.

When evaluating the model fit, the chi-square statistic as well as the CFI, the TLI, and the RMSEA were assessed. Given the sample size (larger than N = 250) and model complexity (involving 30 or more indicator variables), a CFI higher than .90, a TLI higher than .90, and an RMSEA lower than .07 indicate an acceptable fit, while significant p-values are expected for the chi-square test (Hair et al., 2010). This study analyzed the data primarily at the individual level due to the methodological challenges to multilevel modeling, related to changing raters over time while there is only one rating at each occasion (Koch et al., 2020). All the main analyses were conducted using Mplus version 8 (Muthén & Muthén, 1998-2017). The Full Information Maximum Likelihood (FIML) estimator with robust standard errors (MLR in Mplus) was used to handle missingness and any violations to normality and independence assumptions, based on the literature suggesting that Likert-type items possessing 5 or more levels could be reasonably treated as continuous variables (Kline, 2015; Pendergast et al., 2017). With the use of MLR, the chi-square difference test statistics used for invariance testing were calculated employing the Satorra-Bentler scaling correction (Satorra & Bentler, 2010), as suggested by the Mplus developers (see http://www.statmodel.com/chidiff.shtml).

Figure 1.2. Second-order Latent Growth Model of Student Social-Emotional Competence (SEC) Measured by Eight DESSA-Mini Items Across Six Times of Measurement



Results

Descriptive Statistics

The average of 8 DESSA-Mini items, across forms, increased by approximately 0.5 raw score points on a 5-point Likert scale from Year 1 Fall to Year 3 Spring, with an overall tendency of increasing within a year and decreasing over the summer. When transformed into T-Scores based on the national norm (M = 50, SD = 10), this translates into about a 7 T-Score point increase throughout the three years. The distribution of the average scores was slightly left-skewed across all measurement time points, and slightly platykurtic except for Year 3 Spring ratings where it had a positive kurtosis. However, no serious violations to univariate normality were identified as the absolute values of skewness and kurtosis were smaller than 1 for all items, which meets the customary thresholds for skewness and kurtosis (3 and 8, respectively; Kline, 2015). (See Table 1.3).

Table 1.3. Descriptive Statistics of the Items Comprising the DESSA-Mini Forms 1-3

Time	Valid N	Average of 8 items (0-4)				T-Scores	
		M	SD	Skewness	Kurtosis	M	SD
Form 1							
Year 1 Fall	534	2.51	0.97	-0.43	-0.15	47.70	10.07
Year 1 Spring	475	2.90	0.94	-0.57	-0.22	52.96	10.98
Year 2 Fall	633	2.71	0.94	-0.37	-0.24	50.40	11.10
Year 2 Spring	515	2.95	0.96	-0.69	-0.01	54.12	12.15
Year 3 Fall	715	2.82	0.98	-0.49	-0.35	52.30	12.24
Year 3 Spring	395	2.95	0.99	-0.81	0.33	54.27	12.39
Form 2							
Year 1 Fall	534	2.45	1.01	-0.41	-0.23	47.78	10.48
Year 1 Spring	475	2.87	0.96	-0.54	-0.29	53.27	11.20
Year 2 Fall	633	2.67	0.96	-0.36	-0.29	50.67	11.22
Year 2 Spring	515	2.94	0.97	-0.67	-0.07	54.66	12.18
Year 3 Fall	715	2.80	0.98	-0.45	-0.39	52.64	12.20
Year 3 Spring	395	2.93	1.00	-0.78	0.22	54.66	12.41
Form 3							
Year 1 Fall	534	2.37	1.03	-0.34	-0.36	47.42	10.14
Year 1 Spring	475	2.80	1.00	-0.53	-0.39	52.95	11.31
Year 2 Fall	633	2.60	0.99	-0.34	-0.37	50.35	10.94
Year 2 Spring	515	2.90	0.99	-0.64	-0.16	54.50	11.94
Year 3 Fall	715	2.75	1.03	-0.46	-0.45	52.63	12.17
Year 3 Spring	395	2.90	1.03	-0.79	0.22	54.70	12.30

Note. The descriptive statistics were presented at the composite score level due to the space limitations. There was negligible item-level variation in these statistics. The number of valid cases did not differ at the item level at each occasion.

There was a substantial amount of missing data across time. Missing data pattern analysis revealed that 105 students (9.16%) had complete data for all six occasions. About 11% of the students had some valid data for Year 1, but had missing data for all ratings in Year 2 and Year 3. About 13% of the students had some valid data for Year 1 and Year 2, but had missing data for Year 3. Given the high mobility rate of the district, it is likely that many of these cases represent students transferring out of the district during the study years. On the other hand, about 7% of the students had missing data for all Year 1 occasions, but had some valid data for Year 2 and Year 3. About 27% of the students had missing data for all occasions in Year 1 and Year 2, but had some valid data in Year 3. It is likely that many of these cases represent students who were randomly selected to replace students who had left the district. Table 1.4 presents the distribution of missing data patterns.

Additional tests revealed that the baseline SEC scores measured in Year 1 Fall showed a small difference between those who had complete data for all six occasions and those who had some missing data for later occasions, suggesting a higher baseline among students who had complete data ($t = [3.14, 3.44]$, $p < .01$, Hedges' $g = [.30, .33]$ across three forms). The probability of having any missing data did not significantly differ by student gender nor race and ethnicity, but differed by student SES: there were more low-income students among those who had some missing data ($\chi^2(1) = 6.28$, $p = .012$), but the difference was negligible in size (Phi = 0.07). This study applied the FIML approach to missing data when estimating the rate of growth while accounting for the baseline level differences. In addition, the growth modeling results were compared with and without student SES included as a covariate.

Table 1.4. Missing Data Patterns

Missing Pattern	Frequency	Percent
No missing across all three years	105	9.16
Missing earlier ratings		
Missing data for Year 1 + some valid data for Year 2 and Year 3	83	7.24
Missing data for Year 1 and Year 2 + some valid data for Year 3	314	27.40
Missing later ratings		
Some valid data for Year 1 + missing data for Year 1 and Year 2	127	11.08
Some valid data for Year 1 and Year 2 + missing data for Year 1	144	12.57
Other inconsistent missing patterns	373	32.55
Missing all ratings across three years	0	0.00
Total	1,146	100.00

Longitudinal Factorial Invariance Testing Results

To examine the longitudinal factorial invariance of each form of the DESSA-Mini, a series of model fit comparisons with increasing invariance assumptions were conducted. Across forms, although the chi-square difference tests were statistically significant, inappreciable changes were observed in practical fit indices between the configural invariance model and the weak factorial invariance model ($|\Delta CFI| \leq .00$, $|\Delta TLI| \leq .00$, $|\Delta RMSEA| \leq .00$)

as well as between the weak factorial invariance model and the strong factorial invariance model (|ΔCFI| ≤ .01, |ΔTLI| ≤ .01, |ΔRMSEA| ≤ .00). These findings suggest that strong factorial invariance can be assumed, indicating that both factor loadings and intercepts were not different over time. The strong factorial invariance model showed a good fit with each DESSA-Mini Form (CFI ≥ 0.95, TLI ≥ .94, RMSEA ≤ .03). The factor loadings were also high across three forms: Form 1 [.73, .81], Form 2 [.74, .85], and Form 3 [.77, .90]. Although the strict factorial invariance model also showed an acceptable fit, the practical fit indices slightly worsened compared to the strong invariance model (|ΔCFI| ≤ .02, |ΔTLI| ≤ .02, |ΔRMSEA| ≤ .01). As this strict factorial invariance assumption is not a necessary prerequisite for the comparisons of latent means (Widaman et al., 2010), metric and scalar invariance, but not residual invariance, was assumed in the following growth trajectory models. The full invariance testing results are presented in Table 1.5.

Table 1.5. Fit Indices of a Series of Models With Increasing Invariance

Model [Comparison Model]	$\chi^2(df)$ [$\Delta\chi^2(\Delta df)$][a]	CFI [ΔCFI]	TLI [ΔTLI]	RMSEA [ΔRMSEA]
Form 1				
Model 1: Configural invariance	1998.074(945)***	0.96	0.95	0.03
[—]	[—]	[—]	[—]	[—]
Model 2: Weak factorial invariance	2052.96(980)***	0.96	0.95	0.03
[vs. Model 1]	[52.33(35)*]	[.00]	[.00]	[.00]
Model 3: Strong factorial invariance	2194.14(1015)***	0.95	0.95	0.03
[vs. Model 2]	[150.66(35)***]	[-.01]	[.00]	[.00]
Model 4: Strict factorial invariance	2669.85(1055)***	0.93	0.93	0.04
[vs. Model 3]	[302.08(40)***]	[-.02]	[-.02]	[.01]
Form 2				
Model 1: Configural invariance	2119.37(945)***	0.95	0.94	0.03
[—]	[—]	[—]	[—]	[—]
Model 2: Weak factorial invariance	2179.15(980)***	0.95	0.94	0.03
[vs. Model 1]	[58.58(35)**]	[.00]	[.00]	[.00]
Model 3: Strong factorial invariance	2324.54(1015)***	0.95	0.94	0.03
[vs. Model 2]	[154.00(35)***]	[.00]	[.00]	[.00]
Model 4: Strict factorial invariance	2845.46(1055)***	0.93	0.92	0.04
[vs. Model 3]	[327.06(40)***]	[-.02]	[-.02]	[.01]

Table 1.5 (continued).

Form 3				
Model 1: Configural invariance	1604.90(945)***	0.97	0.97	0.03
[—]	[—]	[—]	[—]	[—]
Model 2: Weak factorial invariance	1671.51(980)***	0.97	0.97	0.03
[vs. Model 1]	[67.87(35)**]	[.00]	[.00]	[.00]
Model 3: Strong factorial invariance	1869.08(1015)***	0.97	0.96	0.03
[vs. Model 2]	[213.81(35)***]	[.00]	[-.01]	[.00]
Model 4: Strict factorial invariance	2324.81(1055)***	0.95	0.95	0.03
[vs. Model 3]	[280.32(40)***]	[-.02]	[-.01]	[.00]

[a] The chi-square difference test statistics used the Satorra-Bentler scaling correction.
*p < .05. **p < .01. ***p < .001.

Growth Trajectory Modeling Results

To examine the average SEC growth trajectory over three years, a second-order latent growth model was fitted, assuming strong factorial invariance across time. The model specified as depicted in Figure 1.2 showed a good fit with each DESSA-Mini form: CFI ≥ 0.94, TLI ≥ .94, RMSEA ≤ .03 (see Table 1.6).

Table 1.6. Model Fit of the Second-order SEC Growth Trajectory Model

	$\chi^2(df)$	CFI	TLI	RMSEA	RMSEA 90% CI
Form 1	2297.43(1027)***	0.95	0.94	0.03	[0.031, 0.035]
Form 2	2423.71(1027)***	0.94	0.94	0.03	[0.033, 0.036]
Form 3	1979.66(1027)***	0.96	0.96	0.03	[0.027, 0.030]

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

The mean of the slope factor was estimated to be 0.41 to 0.44, across three different forms, indicating that there was a .41 to.44 standard deviation increase in student SEC from Fall to Spring in Year 1. This change can be interpreted as a small effect size, equivalent to Glass's delta (Glass et al., 1981, as cited in Yang et al., 2020). Combined with the estimated coefficients of the slope factor, the results suggest that there was a decrease in mean SEC during the first summer approximately by .22 standard deviations, about a half of the first-year gain. From Fall to Spring in Year 2, there was an estimated additional .28 to.33 standard deviations increase in mean SEC, which is a small effect size. Over the summer period between Year 2 and Year 3, mean SEC was estimated to decrease again by .15 to .18 standard deviations, about a half of the second-year gain. In Year 3, there was an estimated additional .14 to.15 standard deviations increase in mean SEC from Fall to Spring (See Figure 1.3). In summary, the estimated average SEC growth trajectory was non-linear with an increase within each school year and a decrease over each summer. Estimates of the growth rate within a year and the loss over the summer gradually decreased over time. Table 1.7 summarizes the main findings of the second-order latent growth model.

Table 1.7. Results of the SEC Growth Trajectory Modeling

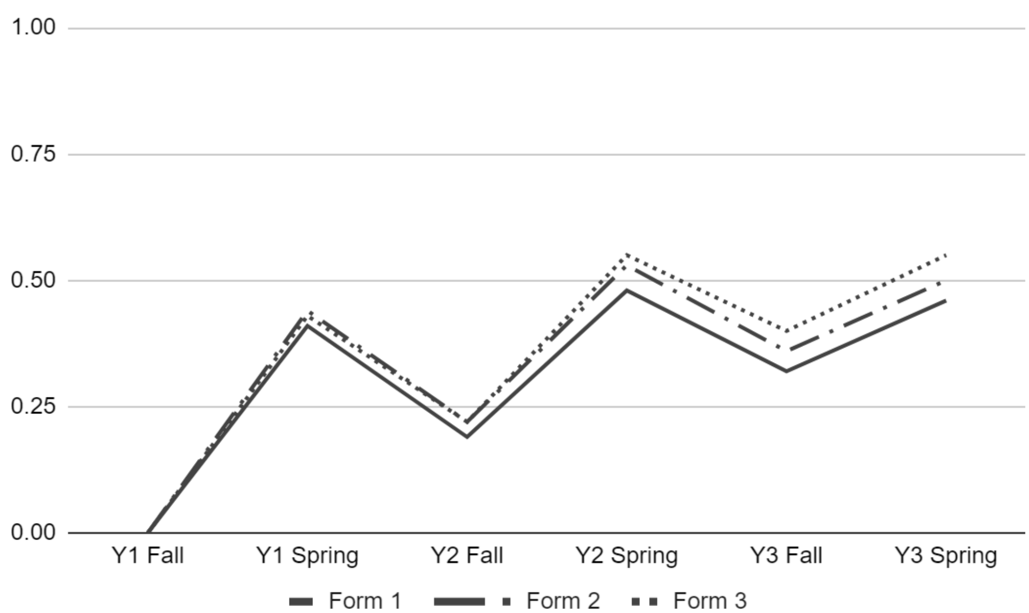
	Form 1			Form 2			Form 3		
	Estimate	SE	*p*	Estimate	SE	*p*	Estimate	SE	*p*
Coefficients of the Intercept factor									
SEC_{Y1Fall}	1.00	—	—	1.00	—	—	1.00	—	—
$SEC_{Y1Spring}$	1.00	—	—	1.00	—	—	1.00	—	—
SEC_{Y2Fall}	1.00	—	—	1.00	—	—	1.00	—	—
$SEC_{Y2Spring}$	1.00	—	—	1.00	—	—	1.00	—	—
SEC_{Y3Fall}	1.00	—	—	1.00	—	—	1.00	—	—
$SEC_{Y3Spring}$	1.00	—	—	1.00	—	—	1.00	—	—
Coefficients of the Slope factor									
SEC_{Y1Fall}	0.00	—	—	0.00	—	—	0.00	—	—
$SEC_{Y1Spring}$	1.00	—	—	1.00	—	—	1.00	—	—
SEC_{Y2Fall}	0.47	0.09	<.001	0.51	0.09	<.001	0.52	0.09	<.001
$SEC_{Y2Spring}$	1.16	0.12	<.001	1.21	0.11	<.001	1.28	0.12	<.001
SEC_{Y3Fall}	0.79	0.11	<.001	0.81	0.10	<.001	0.93	0.10	<.001
$SEC_{Y3Spring}$	1.12	0.13	<.001	1.14	0.12	<.001	1.27	0.13	<.001
Means									
Intercept	0.00	—	—	0.00	—	—	0.00	—	—
Slope	0.41	0.04	<.001	0.44	0.04	<.001	0.43	0.04	<.001
Variances									
Intercept	0.57	0.05	<.001	0.55	0.06	<.001	0.57	0.06	<.001
Slope	0.01	0.00	<.001	0.01	0.00	<.001	0.01	0.00	<.001
Covariances									
Intercept w/ Slope	0.10	0.04	.008	0.08	0.03	.011	0.08	0.03	.005
Residual Variances of SEC Latent Variables									
SEC_{Y1Fall}	0.44	0.05	<.001	0.47	0.05	<.001	0.42	0.05	<.001
$SEC_{Y1Spring}$	0.37	0.05	<.001	0.38	0.04	<.001	0.40	0.04	<.001
SEC_{Y2Fall}	0.46	0.04	<.001	0.44	0.03	<.001	0.43	0.03	<.001
$SEC_{Y2Spring}$	0.46	0.05	<.001	0.45	0.05	<.001	0.39	0.05	<.001
SEC_{Y3Fall}	0.51	0.05	<.001	0.49	0.05	<.001	0.44	0.04	<.001
$SEC_{Y3Spring}$	0.44	0.05	<.001	0.40	0.05	<.001	0.37	0.05	<.001

Table 1.7 (continued).

Estimated Means of SEC Latent Variables									
SEC_{Y1Fall}	0.00	—	—	0.00	—	—	0.00	—	—
$SEC_{Y1Spring}$	0.41	0.04	<.001	0.44	0.04	<.001	0.43	0.04	<.001
SEC_{Y2Fall}	0.19	0.04	<.001	0.22	0.05	<.001	0.22	0.04	<.001
$SEC_{Y2Spring}$	0.48	0.05	<.001	0.53	0.05	<.001	0.55	0.05	<.001
SEC_{Y3Fall}	0.32	0.04	<.001	0.36	0.04	<.001	0.40	0.04	<.001
$SEC_{Y3Spring}$	0.46	0.05	<.001	0.50	0.05	<.001	0.55	0.05	<.001

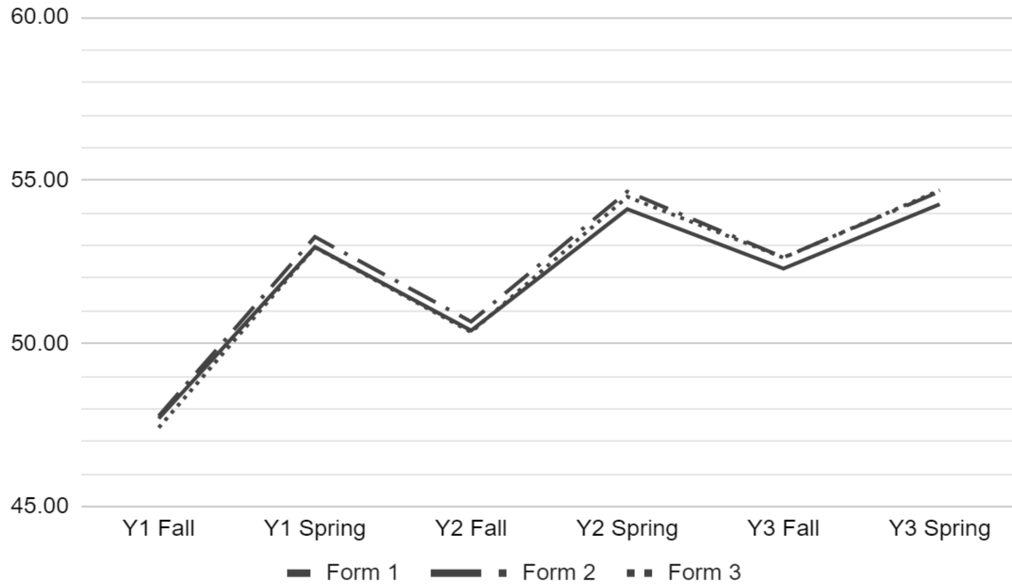
Note. The results of the measurement model assuming strong factorial invariance across time are not presented here for simplicity. The expected values of SEC latent variables in the last six rows were calculated based on the estimated mean of the slope factor and the estimated coefficients of the slope factor. The standard errors of the estimated latent means (reported in the Mplus TECH4 output) were calculated using the delta method.

Figure 1.3. Expected SEC Growth Trajectory Measured by the DESSA-Mini Forms 1-3



Given that most practice settings use the DESSA-Mini composite T-Scores rather than the raw or latent factor scores required for this analysis, the trend of SEC growth using T-transformed scores is also presented in Figure 1.4 to facilitate a more meaningful interpretation. Student SEC in this sample was slightly below the national average (M = 50, SD = 10) at baseline, but increased above it by the end of the first year. Throughout the three years, student SEC increased by approximately 7 T-Score points in total.

Figure 1.4. SEC Growth Trends Using DESSA-Mini Composite T-Scores



Discussion

This study aimed to (1) test longitudinal measurement invariance of a teacher-completed behavioral rating scale for student social-emotional competence (SEC), and (2) examine the average SEC growth trajectory during elementary school years in the context of a school-based universal SEL initiative. Using the DESSA-Mini, a brief behavioral rating scale that has demonstrated sound psychometric properties and is currently in widespread use for SEC progress monitoring in routine practice, this study first examined whether the same construct of SEC was measured consistently across six time points throughout three consecutive elementary school years. The results suggested strong across-time factorial invariance of each of the three DESSA-Mini forms. This finding implies that the same construct of SEC was measured across different seasons within a year (Fall vs. Spring) and across different grade levels, even when mostly rated by different teachers. With this evidence of longitudinal measurement invariance that allows for a valid examination of how student SEC develops over time, this study examined the three-year average growth trajectory of student SEC in the context of a district-wide social and emotional learning (SEL) initiative, using a second-order latent growth model that did not impose any predetermined shape of growth. The results suggest that (a) student SEC increased within each year, (b) student SEC decreased over each summer by about a half of the yearly gain, and (c) the rate of yearly growth gradually decreased across years. These findings contribute to the literature of school-aged children's social-emotional development and school-based SEC assessment, with three key implications for future research and practice.

First, this study demonstrates procedures for testing the measurement invariance of progress monitoring tools over time using longitudinal CFA methods, a well-established statistical modeling approach to testing measurement invariance across time. Although longitudinal measurement invariance is a critical requirement for examining stability and changes of any repeated measures data (e.g., test-retest reliability, pre- and post-test comparisons; Pitts et al., 1996; Widaman et al., 2010), there has been a lack of studies that use advanced longitudinal methods to examine SEC progress monitoring tools. This is the first study that explicitly tested the longitudinal factorial invariance of the DESSA-Mini and, to my knowledge, any of the teacher-completed brief behavioral rating scales currently used for the universal assessment of SEC. The evidence suggesting strong measurement invariance of the DESSA-Mini makes it possible to interpret the observed change in ratings as an actual change in student SEC, rather than reflecting a change in the meaning of SEC or the way it is measured across seasons, grade levels, and raters. This study calls for more studies that appropriately examine longitudinal measurement invariance of SEC progress monitoring tools, in order to accurately measure and quantify social-emotional developmental processes within the school setting.

Second, this study contributes to the existing literature on how elementary school students' SEC develops under routine SEL intervention conditions across school years. Students, on average, showed a statistically significant increase in SEC in the first year (ES = [.41, .44] across forms) as well as in total throughout the three years (ES= [.46, .55] across forms). This growth can be interpreted as a small-to-medium increase per Cohen's criteria (Cohen, 1969), noting that the use of Cohen's standards for educational interventions has been criticized to be too stringent (Kraft, 2019). The observed growth can also be interpreted as a *practically significant* improvement per Ferguson's (2009) criteria for practical significance in social science data (ES > .40). This finding suggests that students in general show socially and emotionally more competent behaviors as they grow in the context of implementing SEL in this racially and ethnically diverse, low-income urban district.

One of the unique findings from this study is the observed non-linear pattern of growth with an increase during a school year and a decrease over the summer. Most of the few existing studies that examined multiyear SEC growth trajectories have assumed linear or curvilinear patterns of growth with no clear theoretical backgrounds. Criticizing these assumptions, Low and colleagues (2019) recently compared the empirical validity of several theorized patterns of two-year SEC growth, including models assuming (a) a consistent linear gain, (b) in-year gains and summer stagnation, (c) in-year gains and summer slide, and (d) Year 1 gain with no further improvement. They found that a model assuming in-year gains with a summer slide was best supported by the data collected from a randomized SEL intervention study for most teacher-rated social-emotional outcomes, including the DESSA. In their best-fitting model, however, the amount of summer loss was fixed to be equal to the previous gains, rather than exploring the pattern of growth, assuming student outcomes reverted to the baseline level after the summer. The current study provides another piece of evidence supporting a pattern of improvement within a year with summer loss, while also suggesting that students might not completely lose all the gains over the summer, but rather show progression across years despite the summer slide. This study, however, did not have the benefit of a comparison group that enables me to tease apart change in response to the SEL program from changes that are a part of natural child development and schooling. More research is called for to expand our knowledge about how student SEC develops within and across school years, and the extent of "social-emotional learning loss" that occurs during the summer or unexpected school closures. Studies with an experimental design would be particularly helpful to understand the factors that contribute to any observed social-emotional learning loss, for example, the discontinuation of SEL interventions and, more broadly, the various social-emotional experiences that schooling provides.

Another noteworthy finding about the SEC growth pattern was that although the slope was positive, the *rate* of growth each year decelerated over time. Some possible explanations underlying this include (a) children show less growth as they grow older, (b) the benefits of an SEL curriculum become smaller as time goes by, (c) the implementation quality of the intervention waned over time. Again, it is beyond the scope of this study to estimate the extent of growth associated with chronological age versus growth in response to a presence or quality of an intervention. Further studies with multiyear SEC progress monitoring data are called for to understand whether student SEC growth trajectory shows a similar decelerated pattern of growth across different samples and contexts, and to investigate its underlying factors, ideally in an experimental setting where implementation quality is also monitored across time.

For practitioners, the findings of this study could imply that students might show a small increase in SEC in the first year, and medium-size increase throughout two or three years of the routine implementation of an evidence-based SEL curriculum, despite some summer slides, even with Cohen's stringent criteria applied. It is also important to note that in many schools, student SEC is monitored by rotating forms (i.e., presenting 8 distinct items for each monitoring period) within a year to avoid practice effects. While it is not possible, methodologically, to test the measurement invariance across forms composed of non-overlapping items, this study suggests this to be a reasonable practice, given very high alternate form reliability estimates (as described in the methods section), and negligible across-form level differences at each occasion as shown in Figures 1.3 and 1.4.

Despite the familiar limits of the generalizability of these projections to other samples, settings, interventions, and time, and several limitations related to methodological constraints (e.g., high missingness due to students' in-and-outs, certain subgroups excluded due to small group size, classroom- or school-level nestedness not accounted for due to changing raters), this study provides a useful piece of evidence for how school-aged children's SEC develops under a routine SEL implementation condition, *on average*. Yet, a significant amount of variability in this average SEC growth trajectory was observed, especially in the baseline level. This finding raises an important question of what explains these variations, who shows more or less favorable trajectories, and whether there exist any sociocultural and structural disparities in these trajectories when there is no reason to believe that student SEC can vary by certain demographic characteristics themselves. This study calls for future research that prioritizes the examination of subgroup disparities in the development of SEC.

Chapter 2

Gender, Racial-Ethnic, and Socioeconomic Disparities in the Development of Social-Emotional Competence Among Elementary School Students

---

Summary

Social-emotional competence (SEC) has been demonstrated to be a crucial factor for student mental health, malleable through the high-quality implementation of effective school-based social and emotional learning (SEL). Although SEL is now widely practiced in the U.S. as a Tier-1 intervention for the entire student body, it remains unclear whether disparities exist in the development of SEC across socioculturally defined subgroups of students. Also, despite the field's widespread concern about teacher bias in assessing SEC within diverse student bodies, little evidence is available on the measurement invariance of the SEC progress monitoring tools used to explore and facilitate SEC development. Based on a sociocultural view on student SEC development, this chapter presents a study that aims to appropriately measure and examine the extent to which gender, racial-ethnic, and socioeconomic disparities exist in SEC developmental trajectories during elementary school years. Specifically, using three years of SEC progress monitoring data collected from a districtwide SEL initiative (N = 5,452; Grades K-2 at baseline; 9 times of assessment), this study (1) tested the measurement invariance of a widely-used, teacher-rated SEC assessment tool (DESSA-Mini) across student gender, race and ethnicity, and socioeconomic status (SES), and (2) examined the extent to which multiyear SEC growth trajectories differed across these subgroups under a routine SEL practice condition. The invariance testing results supported strict factorial invariance of the DESSA-Mini across all the examined subgroups, providing a foundation for valid cross-group comparisons of student SEC growth. The piecewise latent growth modeling results indicated that male (vs. female), Black or Hispanic (vs. White), and low-income (vs. middle-to-high-income) students started with a lower level of SEC, and these gaps were either sustained or slightly widened throughout three elementary school years. Based on these findings, this study calls for future research that can inform practice efforts to ensure equitable SEL assessments and produce more equitable SEL outcomes, thereby promoting equity in school mental health.

Background

Social-emotional competence (SEC) has been defined as the knowledge, attitudes, and skills necessary to know and manage the self, understand and relate to others, and make responsible personal and social decisions (Elias et al., 1997; Weissberg et al., 2015). Children's SEC is found to be a reliable protective and promotive factor of mental health and many other positive developmental outcomes, including behavioral health and educational success

(Denham, 2006; Domitrovich et al., 2017; Zins et al., 2004). The development of SEC occurs at various developmental stages, through various socialization contexts: In infancy and toddlerhood, children typically develop a sense of self and others, emotional and behavioral control, prosocial emotion and behavior, as well as an understanding of social conventions and norms (Shulman, 2016; Wittmer et al., 1996). In this early stage of life, the family environment plays a critical role in children's social and emotional development (Shulman, 2016). In middle childhood, school becomes another important socialization context. School entry (i.e., starting kindergarten in the U.S. context) provides a new set of opportunities and challenges for children to develop SEC, as they enter into a distinctive social world from earlier socialization experiences and spend a significant amount of time there (W. A. Collins, 1984). However, individual children enter schools with different levels of SEC, and the variability in SEC is not random across diverse student subgroups (Raver & Knitzer, 2002; Yates et al., 2008).

Gender, race and ethnicity, and socioeconomic status (SES) have been conceptualized as important factors associated with individual variability in SEC that manifest in student behaviors in the school context (Garner et al., 2014). These characteristics are attributed to individuals, but the ways in which these developmental differences in SEC emerge involve the influence of various environmental and structural factors that are differentially experienced by diverse subgroups. From a sociocultural perspective, Garner and colleagues (2014) provide a heuristic model depicting the interrelations between the individual characteristics and the social-emotional practices of their family, community, and school contexts, as related to the development of student SEC. In their explanations of how gender, race and ethnicity, and SES can each be linked to SEC development, the authors allude to at least two mechanisms, one related to different socialization processes and the other related to structural inequalities.

A first explanation postulates that children of diverse genders, races, ethnicities, and socio-economic classes may experience different socialization processes, oriented to the emotional and behavioral norms and expectations specific to the subgroup, which may or may not be aligned with the dominant norms and expectations of schools. In this view, some students' behaviors, manifested in ways that are *consistent* with the sociocultural norms and expectations of the subgroup, may be *inconsistent* with the norms and expectations of the school system. Thus, these children may be assessed to have less SEC by teachers in schools. In the American education system, which is emergent from White, middle-class values and norms (Boykin et al., 2005; Leonardo, 2012), sustained by a mostly White, female, professionalized workforce (Institute of Education Sciences, 2020), certain behaviors more frequently shown among male students, students of color, and lower-class students could be viewed as less socially and emotionally competent than those of female, White, and upper-class students.

A second explanation postulates that children of diverse races, ethnicities, and socio-economic classes may experience different sets of opportunities and barriers that differentially promote or hinder their SEC development due to structural inequalities. In this view, students from marginalized backgrounds may be deprived of various opportunities and resources to develop SEC, relative to students from advantaged backgrounds. The history and present realities of racism and classism in the U.S. have created different developmental environments for White children versus children of color, and for upper-class children versus lower-class children, including parents' working conditions, neighborhood environments, and interpersonal experiences at schools and other public spaces (Dixon-Román, 2017; Lareau, 2011; Leonardo, 2004). Students of color and lower-class students are also at greater risk of encountering implicit bias, microaggressions, overt discrimination, and victimization, which all have adverse effects on their social-emotional development (Sarcedo et al., 2015; Sue, 2010; Zoric, 2014). In this context, racially, ethnically, and socioeconomically marginalized students may develop lower SEC compared to students from advantaged backgrounds.

These two potential mechanisms, which are by no means mutually exclusive, together provide a useful conceptual framework to understand why and how gender, racial-ethnic, and socioeconomic disparities in student SEC may exist in the U.S. context. Empirical evidence, however, has been limited and inconsistent as to the extent to which and how disparities in SEC manifest and how they develop throughout the school years. Using this sociocultural view on student SEC development as a guiding framework for interpretation and discussion, the current study attempts to accurately measure and understand the extent to which gender, racial-ethnic, and socioeconomic disparities in SEC developmental trajectories exist among school-aged children in the U.S. This first phase of *detecting* disparities will set the stage for the next phases of research on *understanding* and *reducing* disparities in student social-emotional outcomes, in hope of promoting equity in school mental health (Kilbourne et al., 2006).

Empirical Evidence on Disparities in Social-Emotional Competence Development

Overall, the literature lacks studies whose main objective is to investigate subgroup disparities in SEC developmental trajectories among school-aged children, particularly when compared to ample evidence on disparities in overall mental, emotional, behavioral health problems (e.g., Alegría et al., 2015). Existing research on subgroup differences in social and emotional development has mainly focused on gender differences (Eisenberg et al., 2006; Zahn-Waxler & Smith, 1992), but evidence is scarce as to differences associated with other sociocultural subgroups, and even more so as to how these subgroup differences unfold over the course of school years. The literature on school-based social and emotional learning (SEL) interventions may provide insights into these questions, as a program evaluation study typically involves repeated measures of SEC within or across school years among entire student populations that may consist of diverse student subgroups. Despite a growing body of SEL intervention research, however, many of these studies have not even described their samples in regard to student gender, race and ethnicity, or SES. According to a review of 117 peer-reviewed articles included in Durlak and colleagues' (2011) influential meta-analysis of SEL program effectiveness, 31% did not provide sample descriptions for gender, 36% for race and ethnicity, and 55% for SES (Rowe & Trickett, 2018).

In the available literature, including both observational and intervention research, the most commonly reported subgroup difference in SEC involves gender. Kindergarten and elementary school boys have been observed to have lower SEC than girls across diverse measurement methods such as teacher ratings (Aber et al., 2003; Bierman et al., 2010; Frey, Hirschstein, et al., 2005; Hutchison et al., 2020; S. M. Jones et al., 2011; Krishnan, 2011; Malti et al., 2011), self-reports (Holsen et al., 2008; Malti et al., 2011; West et al., 2020), peer nominations (Bierman et al., 2010), and task-based assessments (Colle & Del Giudice, 2011). Yet, findings about gender differences in the rate of SEC growth are quite rare and mixed. In the context of SEL intervention research, boys showed less growth than girls through teacher ratings of prosocial behavior in one study (Aber et al., 2003), but they had a similar rate of growth to girls in teacher ratings of prosocial behavior and emotion regulation in another study (S. M. Jones et al., 2011).

Evidence regarding racial-ethnic differences in SEC is more limited in volume and more inconclusive. In a large-scale observational study of students in Grade 4 to 12, African American, Asian, and Latinx students rated themselves, on average, to have lower SEC than White peers (West et al., 2020). Some SEL intervention studies using teacher ratings also found that Black/African American students had lower SEC scores at baseline compared to White students (Aber et al., 2003; S. M. Jones et al., 2011) or to non-Black students (Elias & Haynes, 2008), and that Hispanic students had lower baseline SEC than non-Hispanic White students (Aber et al., 2003). Other studies, however, found that teacher ratings of SEC did not differ by race (Chain et al., 2017) or ethnicity (S. M. Jones et al., 2011). With regard to racial-ethnic differences in the rate of growth, in the context of SEL intervention research, some studies have found no differences in the rate of growth by race and ethnicity (Elias & Haynes, 2008; S. M. Jones et al., 2011), while one found that Black students showed a slower increase than White peers as assessed through teacher ratings of prosocial behavior (Aber et al., 2003).

Findings on the differences in SEC by family SES are also limited in volume and inconsistent in their findings. Economically disadvantaged students had lower SEC than other students in a large-scale self-report survey (West et al., 2020). Some SEL intervention studies have similarly reported lower levels of SEC among students of lower SES, as assessed by teacher ratings (Aber et al., 2003) or child self-reports (Holsen et al., 2009), but another SEL intervention study found no baseline differences in SEC using teacher ratings (S. M. Jones et al., 2011). In one study that used a multi-modal assessment strategy, teacher ratings indicated lower prosocial behavior among lower SES students, but no such difference was observed with student self-reports (Malti et al., 2011). With regard to socioeconomic subgroup differences in SEC growth rates in SEL intervention research, available evidence suggests that the rate of SEC growth trajectories, as rated by teachers, did not differ by student SES (Aber et al., 2003; S. M. Jones et al., 2011).

In summary, evidence is inconclusive as to the extent to which gender, racial-ethnic, and socioeconomic disparities in student SEC development exist and develop over the course of time. Yet, one could tentatively say that (a) more consistent findings have been reported on gender differences in the overall level of SEC, and (b) when significant subgroup differences were observed, the results seem to suggest less favorable SEC growth trajectories for males, students of color, and low-SES students.

These observed differences, however, can only be interpreted as real subgroup disparities in SEC, if one assumes that the measures of SEC are not biased in their assessment of specific subgroups. For example, some scholars have questioned whether the gender differences in prosocial behavior found in the literature may be due, in part, to biases in measures (e.g., consisting of more “feminine” than “masculine” items; Eisenberg et al., 2006; Zahn-Waxler & Smith, 1992; Zarbatany et al., 1985). But little evidence is available on the subgroup comparability of the measures used in the studies reviewed above. In the following section, the issue of measurement bias and the importance of testing measurement invariance are further discussed. Along with the evidence of measurement invariance, the literature needs more studies exploring subgroup disparities in student SEC development which will ultimately advance our knowledge of the mechanisms underlying any existing disparities and the impact of interventions on these gaps.

Examining SEC Disparities Using Teacher-completed Behavioral Rating Scales: The Importance of Testing Measurement Invariance

One critical prerequisite for examining subgroup disparities in SEC is to make sure that the SEC measurement tool in use is comparably valid, or measures the same construct, across subgroups (Meredith & Teresi, 2006; Putnick & Bornstein, 2016). Measurement bias can have different meanings in various contexts, but this study focuses on a statistical phenomenon where observed scores are systematically influenced by factors other than the actual amount of the construct being measured. With respect to cross-group comparisons, a measure is said to be biased if it functions in a systematically different way across subgroups, making an accurate assessment and comparison impossible (Millsap & Everson, 1993).

With an increasingly widespread adoption of school-based SEL initiatives across the U.S., researchers and practitioners have described the need for SEL assessment tools that can be equitably applied across diverse student populations (Assessment Work Group, 2019; Garner et al., 2014; Mahoney et al., 2022). In elementary education, teacher-completed behavioral rating scales have become the most commonly used tools for universal screening and progress monitoring in the domain of SEL. Although no assessment method is free from the risk of measurement bias, many have questioned the cross-group validity of teacher behavioral ratings, given the persistent history of the overrepresentation of marginalized students in school discipline and special education programs (Peters et al., 2014; Skiba et al., 2011; Zhang & Katsiyannis, 2002). With a growing critique of SEL being colorblind and reflecting white, middle-class values (Gregory & Fergus, 2017; Hoffman, 2009; Mahfouz & Anthony-Stevens, 2020), there is an emerging consensus that we must use SEL measures that are relevant and fair across diverse student populations in order to avoid reinforcement or reproduction of structural inequalities (Assessment Work Group, 2019).

The potential risk of measurement bias associated with the assessment of diverse student subgroups with teacher behavioral ratings implores us to empirically examine the extent to which such bias exists. Alternatives to the empirical interrogation of bias, such as just assuming that the measure is unbiased (and therefore using it recklessly) or just assuming the measure is biased (and advocating against accurately understanding the presence of disparities and addressing them) are each unacceptable. A test of measurement invariance (also known as measurement equivalence) across student subgroups has been proposed as one way to quantitatively explore measurement bias in SEL assessments (Assessment Work Group, 2019; Gehlbach & Hough, 2018). Measurement invariance means that the relation between observed scores and a latent construct is the same across different groups, while measurement non-invariance means that the construct measured has a different structure or meaning as applied to different groups (Putnick & Bornstein, 2016). Although evidence of measurement invariance does not in and of itself guarantee that a measure is bias-free, it is one of the first necessary steps to ensure that a measure is relevant and fairly applied across diverse subgroups (Pendergast et al., 2017).

Research has demonstrated that measurement non-invariance can lead to erroneous conclusions about cross-group comparisons, including the subgroup differences in growth trajectories—both in understanding the initial level and the rate of growth (Chen, 2008; E. S. Kim & Willson, 2014). Despite this risk, only a few recent studies have explicitly examined the measurement invariance of SEC assessment tools. Some self-report instruments have demonstrated invariance across gender, race and ethnicity, and/or SES (Anderson-Butcher et al., 2016; Gehlbach & Hough, 2018; T. M. Jones, 2018; Mantz et al., 2018; McKown, 2019b), and some teacher-completed measures have demonstrated invariance between female and male (Fredrick et al., 2019; Rodkin et al., 2013) and between Black and White students (Pendergast et al., 2017; Rodkin et al., 2013). Yet, these measures are not necessarily the ones that are widely used in research and practice, informing our understanding of subgroup differences as they are enacted upon in schools. Much more work is needed to establish evidence of measurement invariance with diverse student populations, and ultimately to advance our knowledge base on gender, racial-ethnic, and socioeconomic disparities in student SEC development.

Research Objectives

The objective of this study is twofold: (1) to test measurement invariance of a widely used teacher-completed behavioral rating scale for SEC (Devereux Student Strengths Assessment Mini [DESSA-Mini]; Naglieri et al., 2011/2014) across student gender, race and ethnicity, and SES, and (2) to examine the extent to which student SEC growth trajectories differ across these student subgroups during the elementary school years in the context of a district-wide SEL intervention.

Using available school administrative data, this study is primarily focused on comparisons between male and female students; Black/African American ("Black"), Hispanic/Latinx of any race ("Hispanic"), and non-Hispanic White ("White") students; and low-income students eligible for free or reduced-price meals and middle-to-high-income students not eligible for free or reduced-price meals. I recognize that these aspects of individual identities are socially constructed, and the above categorizations may not accurately represent how students identify themselves, and can obscure a great deal of within-group heterogeneity. Acknowledging these limitations, I used available data to illuminate similarities and differences in SEC development. To be clear, this study does not test differential intervention effects across subgroups in an experimental sense. Rather, it compares student SEC growth trajectories across subgroups under a single, routine SEL implementation condition.

Methods

Data and Sample

Data used in this study come from a district-wide SEL initiative that implemented the Promoting Alternative Thinking Strategies (PATHS) curriculum (Kusché & Greenberg, 1994) in a large urban district for three academic years from 2011-2012 to 2013-2014. This district serves a racially and ethnically diverse and low-income student body in Pennsylvania, characterized by a high mobility rate. During the study years, the district's average elementary school enrollment was 8823, 64.5% of whom were identified as Hispanic of any race, 14.8% non-Hispanic Black, and 13.2% non-Hispanic White, with approximately 86% receiving free or reduced-price meals and 42% transferring in and out of schools (Pennsylvania Department of Education, 2021). The PATHS curriculum is a classroom-based universal SEL program found to be effective in promoting social-emotional, behavioral, and academic outcomes among school-aged children in several randomized trials (e.g., Conduct Problems Prevention Research Group, 1999; Fishbein et al., 2016; Greenberg et al., 1995; Schonfeld et al., 2015). The PATHS curriculum was implemented with K-2 students across 15 elementary schools in Year 1 (2011-2012), and with all K-5 students across these 15 elementary schools in Year 2 (2012-2013) and Year 3 (2013-2014), except three schools that elected to discontinue implementing PATHS starting Year 2. Teachers were provided with initial and booster training sessions, along with ongoing support and classroom visits from coaches providing technical assistance. Program implementation in this larger project is described elsewhere (e.g., Fleming, 2014; Shapiro et al., 2018). Study protocols were approved by the Institute of Clinical Training and Research at the Devereux Foundation. De-identified data (no longer regarded as human subjects data) was subsequently shared with the first author, whose analytic activities were acknowledged by the Committee on Protection of Human Subjects at the University of California, Berkeley.

The sample of this current study consists of all students who were in grades K through 2 in Year 1 (grades 2-4 in Year 3; N = 5,452), who were exposed to the PATHS implementation starting Year 1. In this sample, grade and gender were quite evenly distributed: 34.37% Kindergartners, 33.07% first graders, and 32.56% second graders at baseline; 47.95% female and 52.02% male. The sample was representative of the district's student population in terms of the distribution of race and ethnicity as well as SES indicated by free or reduced-price meals eligibility: 63.92% identified as Hispanic, 15.79% Black, 13.65% White, and 87.07% low-income. Most students (77.71%) were assessed exclusively while affiliated with PATHS-implementing schools, while the rest of the students were assessed, at some point in Year 2 and Year 3, while affiliated with one of the schools that discontinued the implementation of PATHS starting Year 2. A very small number of students (less than 0.2%) who had missing information on gender, race and ethnicity, or SES were excluded from analysis. Table 2.1 shows the sociodemographic characteristics of the study sample.

Table 2.1. Sample Characteristics

Variable	Category	Frequency	Percent
Grade at Baseline	K	1,874	34.37
	1	1,803	33.07
	2	1,775	32.56
Gender	Female	2,614	47.95
	Male	2,836	52.02
	Missing	2	0.04
Race and Ethnicity	American Indian or Alaska Native	9	0.17
	Asian	67	1.23
	Black (non-Hispanic)	861	15.79
	Hawaiian or Pacific Islander	9	0.17
	Hispanic (of any race)	3,485	63.92
	White (non-Hispanic)	744	13.65
	Multi-race	275	5.04
	Missing	2	0.04
Socioeconomic Status	Low-income	4,747	87.07
	Middle-to-high income	697	12.78
	Missing	8	0.15
Exposure to the PATHS implementation	Continued for three years	4,237	77.71
	Discontinued in Years 2-3	1,215	22.29
Total		5,452	100.00

Measurement

Student SEC was repeatedly measured using the DESSA-Mini (Naglieri et al., 2011/2014) three times a year for three sequential school years. Fall ratings were collected using the DESSA-Mini Form 1 around October, Winter ratings using Form 2 around January, and Spring ratings using Form 3 around June. The DESSA-Mini, an 8-item version of the full 72-item DESSA (LeBuffe et al., 2009/2014), is a brief, standardized, and norm-referenced behavioral rating scale that assesses K-8 students' overall level of SEC (Naglieri et al., 2013). Each DESSA-Mini form is comprised of 8 non-overlapping items from the full DESSA. Teachers rate the frequency of a student's behavior observed during the past four weeks on a 5-point Likert scale (0=Never, 1=Rarely, 2=Occasionally, 3=Frequently, 4=Very frequently). All of the items are strengths-based, indicating positive rather than maladaptive behaviors (Simmons et al., 2016). Example items include "keep trying when unsuccessful," "respect another person's opinion," "do something nice for somebody," and "show good judgment." In practice, item raw scores are transformed into T-Scores based on a nationally representative standardization sample (M = 50, SD = 10) to be used to assess individual students' relative social-emotional strengths and needs. T-Scores of 60 and above indicate that the student has *Strengths*, T-Scores between 41 and 59 represent *Typical* SEC, and T-Scores of 40 and below indicate a *Need for Instruction*.

The DESSA-Mini has demonstrated strong psychometric qualities including reliability (e.g., internal reliability, test-retest reliability, inter-rater reliability, alternate form reliability) and validity (e.g., concurrent and predictive criterion validity), as evidenced in several studies (Lee et al., 2018; Shapiro, Accomazzo, et al., 2017; Shapiro, Kim, et al., 2017). In the current study, the internal reliability coefficient estimates (Cronbach's alpha) range from .95 to .98 across nine times of measurement, which were slightly higher than those reported in the original development study (.919-.924; Naglieri et al., 2011/2014). Evidence of longitudinal measurement invariance of each form (i.e., strong factorial invariance across time) and interchangeability across forms (i.e., alternate form reliability coefficients ranging from .95 to .98) has been demonstrated in Lee et al. (2022b). Longitudinal measurement invariance within and across years was also confirmed with the current study sample, in which over 90% of the students were rated by the same teachers within a single year, and over 96% by different teachers across years.

Student sociodemographic characteristics were collected from an administrative database. In the original source of data, student gender was measured as a binary variable (female or male), race and ethnicity was measured as having seven categories as presented in Table 2.1, and SES was indicated by whether or not the student was eligible for free or reduced-price meals, determined as a function of family annual income and federal poverty guidelines (e.g., the annual income eligibility criteria in 2011-2012 was $41,348 for a household size of four). In the analyses that involved the race and ethnicity variable, I had to limit my sample to the three largest subgroups (Black, Hispanic, and White) that together comprised the majority (93.4%) of the sample, because the inclusion of the other subgroups in measurement invariance testing, either each as a unique subgroup or all together as a fourth group, yielded model convergence problems likely due to small group sizes. When comparing subgroup differences in SEC growth trajectories using conditional growth trajectory modeling, student gender and SES were dummy coded in ways that female and middle-to-high income subgroups, respectively, had the value of 0, while male and low-income subgroups, respectively, had the value of 1. Implications and limitations of these measurement approaches are discussed later.

Analysis Plan

Factorial Invariance Testing Across Subgroups.

To test measurement invariance of the DESSA-Mini across student subgroups, this study conducted multigroup confirmatory factor analysis (MG-CFA), a structural equation modeling approach to measurement invariance tests. MG-CFA is one of the most commonly used approaches in educational and psychological research and has been demonstrated to have relative strengths to other methods of testing measurement invariance, such as item response theory (IRT) or multiple indicator multiple cause (MIMIC) modeling (see Pendergast et al., 2017 for a review). One of the most important advantages of using MG-CFA, specific to this study, is that it can be directly incorporated into second-order latent growth modeling to compare SEC growth trajectories across subgroups (E. S. Kim & Willson, 2014).

Using this approach, four levels of factorial invariance (i.e., configural invariance, weak factorial invariance, strong factorial invariance, and strict factorial invariance) were tested across student gender (female vs. male), race and ethnicity (Black vs. Hispanic vs. White), and SES (low-income vs. middle-to-high income). In MG-CFA, the *configural* invariance model specifies the identical configuration or structure of the factor-indicator relations across subgroups, while placing only minimal constraints for model identification. Evidence of configural *non*-invariance suggests that the structural relations between each item and the SEC construct is different across subgroups and no subgroup comparisons should be made. Configural invariance is a prerequisite for all the following invariance tests.

The *weak* factorial invariance model adds cross-group equality constraints on the factor loading for the same indicator variable (i.e., metric invariance) to the configural invariance model. This relates to the magnitude of the relation between each item and the SEC construct. Evidence of metric *non*-invariance suggests that teacher ratings of certain items are a better representation of the overall SEC for one group than another and the model should be refined. Configural and metric invariance is a prerequisite for further invariance tests, and without further invariance tests, the measure is considered to have weak factorial invariance, which does not allow valid comparisons of factor means across subgroups.

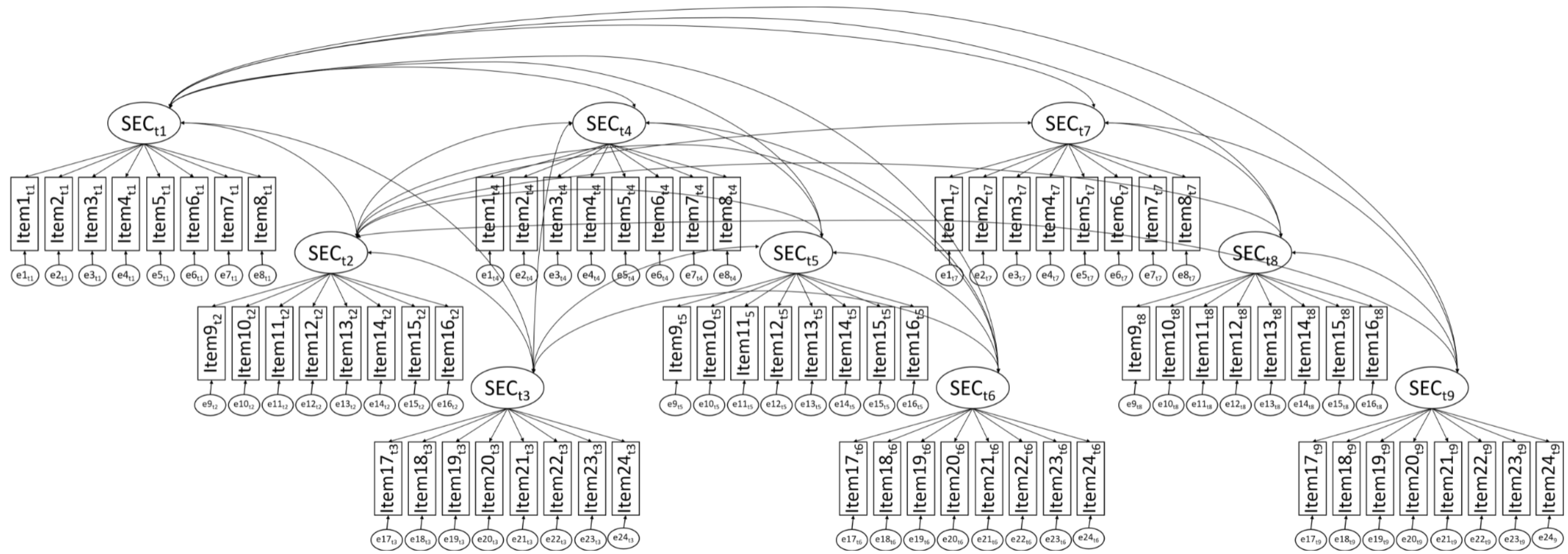
The *strong* factorial invariance model adds cross-group equality constraints on the intercept of the same indicator variables (i.e., scalar invariance) to the weak factorial invariance model. This relates to the intercept, or a mean of each item compared to the mean of the construct. Evidence of scalar *non*-invariance suggests that teachers rate certain items systematically lower or higher for one group than another, independent of students' true SEC level. Scalar invariance, in addition to configural and metric invariance, implies strong factorial invariance, allowing cross-group comparisons of factor means, which is a prerequisite condition for examining subgroup disparities in growth trajectories.

Finally, the *strict* factorial invariance model adds cross-group equality constraints on unique variance of the same indicator variables (i.e., residual invariance) to the strong factorial invariance model. This relates to the residual variance of each item, or the item-unique variance remained unexplained by the SEC construct. Residual invariance, in addition to configural, metric, and scalar invariance, implies strict or full factorial invariance. As the residuals are not part of the latent construct, however, many researchers suggest that residual non-invariance is inconsequential to interpreting factor mean differences (Pendergast et al., 2017; Putnick & Bornstein, 2016).

The longitudinal confirmatory factor model to be compared across subgroups is presented in Figure 2.1. This model consists of a total of 24 indicator variables comprising three forms of the DESSA-Mini (Item1 to Item8 in Form 1, Item9 to Item16 in Form 2, and Item17 to Item24 in Form 3). The eight indicator variables comprising each form are hypothesized to measure a single latent variable (SEC) at each of nine times of measurement (t1 to t3 in Year 1, t4 to t6 in Year 2, and t7 to t9 in Year 3). The SEC latent variables were allowed to covary across time as indicated by double-headed, curved arrows in Figure 2.1. The residuals for the same indicator (e1 to e24) were also allowed to covary across occasion although not shown in Figure 2.1 for graphical simplicity. In this model, strong longitudinal factorial invariance was first tested and confirmed by constraining the factor loadings and intercepts of the same indicator variable to be equal across occasions.

The four levels of invariance tests involve model fit comparisons of a series of nested models with increasing restrictions. Several model fit indices were compared to examine the degree to which these equality assumptions held with data. The chi-square difference test based on the likelihood ratio test statistics was used to test the statistical significance of the model fit differences. Since the chi-square statistic is very sensitive to sample size, however, alternative fit indices including the comparative fit index (CFI), the Tucker-Lewis index (TLI), and the root mean squared error of approximation (RMSEA) were also used. Given the large sample size of the current study, when the chi-square difference test suggested a significantly worse fit for a more restrictive model, but no appreciable change was found in alternative fit indices (e.g., ΔCFI smaller than or equal to $|0.01|$; Cheung & Rensvold, 2002), the more restricted model was assessed acceptable.

Figure 2.1. Longitudinal Confirmatory Factor Model of Student Social-Emotional Competence (SEC) Measured by Three DESSA-Mini Forms Across Nine Times of Measurement



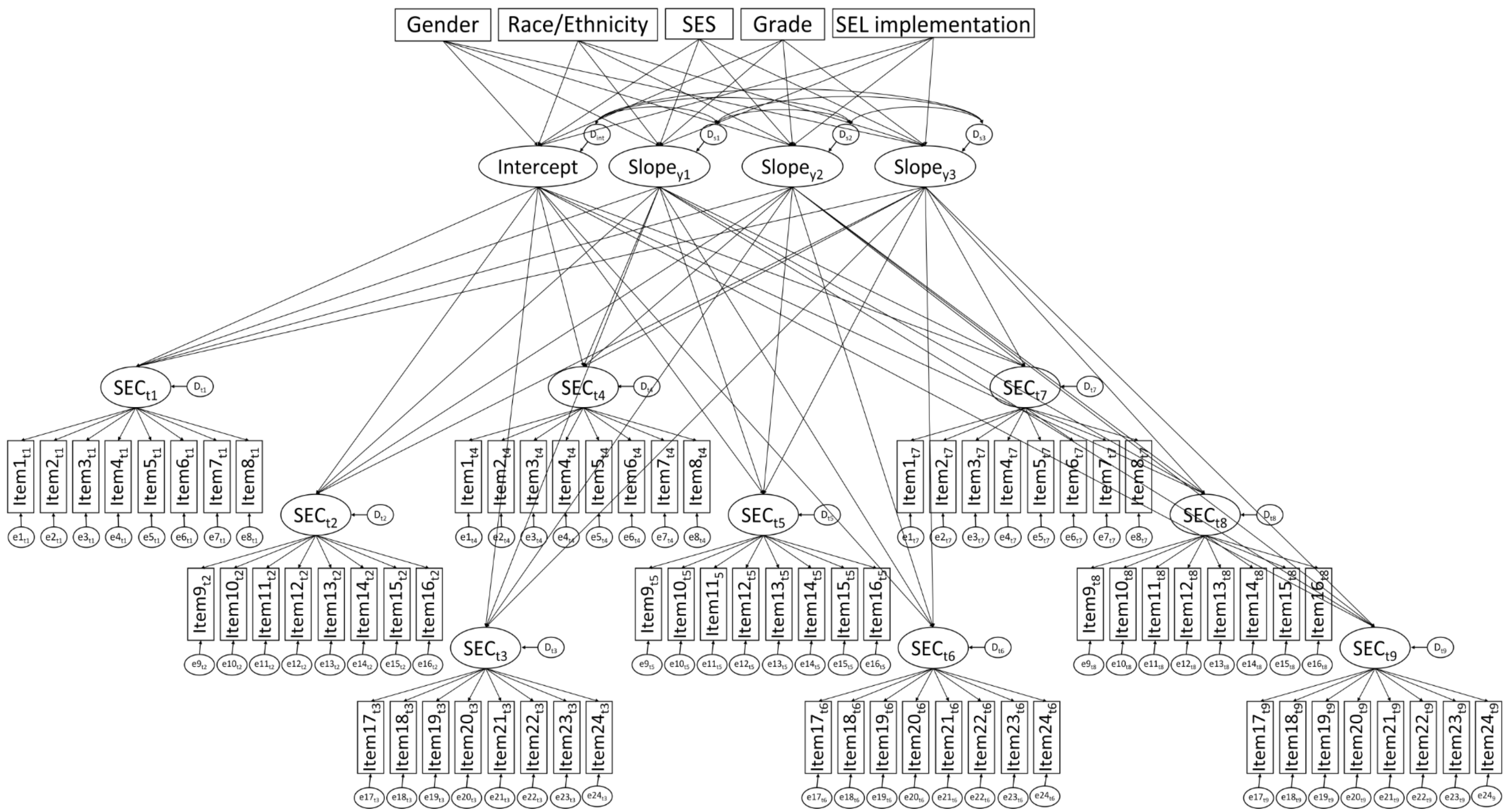
Growth Trajectory Modeling.

After confirming the comparability of latent means across subgroups with at least strong factorial invariance, the growth trajectories of student SEC were estimated using piecewise second-order latent growth modeling in order to incorporate separate growth profiles for each academic year, while simultaneously fitting a measurement model. The model analyzed to examine subgroup differences in student SEC growth trajectories is presented in Figure 2.2.

The first-order component of this model is identical to the measurement model shown in Figure 2.1, assuming metric and scalar invariance across time and with residuals for the same indicators allowed to covary across time. The second-order component of the growth model consists of four growth factors, intercept (baseline level) and three separate slopes (growth rates) corresponding to each academic year ($Slope_{Y1}$ to $Slope_{Y3}$). Instead of assuming a linear or other predetermined shape of the growth curve within each year, the coefficients in the middle of the corresponding year's slope factor were freely estimated, while the first and third coefficients of the corresponding year's slope factor were fixed at 0 and 1, respectively. In order to incorporate the findings of the study presented in Chapter 1 on summer slides to be about a half of the previous year's growth, the 4^{th} to 9^{th} coefficients of the Year 1 slope as well as the 7^{th} to 9^{th} of the Year 2 slope were fixed at 0.5. With the latent standardization identification method applied as suggested by Yang et al. (2020), the mean of the intercept factor is fixed at 0, and the mean of each slope factor indicates the mean change scores between Fall and Spring within the corresponding year relative to the standard deviation of SEC scores at baseline.

The fit of the unconditional growth model was first assessed before including any covariates of the SEC growth trajectory, based on the chi-square statistic as well as the CFI, the TLI, and the RMSEA. Given the sample size (larger than N = 250) and model complexity (involving 30 or more indicator variables), a CFI higher than .90, a TLI higher than .90, and an RMSEA lower than .07 indicate an acceptable fit, while significant p-values are expected for the chi-square test (Hair et al., 2010, p. 20). After confirming the goodness of fit of the unconditional growth model, student gender, race and ethnicity, and SES were included as a set of predictors of the intercept and three distinct slope factors, along with two other covariates: student grade level (0 = grade K, 1 = first grade, 2 = second grade) and the exposure to continued PATHS implementation for three years (0 = discontinued in Year 2 and Year 3, 1 = continued for three years). This analysis was conducted with both the full sample and the subsample that included only three racial-ethnic subgroups (Black, Hispanic, and White) as used in the tests of factorial invariance by race and ethnicity. Although no notable difference was found in the model fit indices and parameter estimates between two samples, I present and interpret the findings with the subsample including only three racial-ethnic subgroups, with which factorial invariance has been tested and established. This study analyzed the data primarily at the individual level due to the methodological challenges to multilevel modeling, related to changing raters over time while there is only one rating at each occasion (Koch et al., 2020). All the main analyses were conducted using Mplus version 8 (Muthén & Muthén, 1998–2017). The Full Information Maximum Likelihood (FIML) estimator with robust standard errors (MLR in Mplus) was used to handle missingness and any violations to normality and independence assumptions, based on the literature suggesting that Likert-type items possessing 5 or more levels could be reasonably treated as a continuous variable (Kline, 2015; Pendergast et al., 2017). With the use of MLR, the chi-square difference test statistics used for invariance testing were calculated employing the Satorra-Bentler scaling correction (Satorra & Bentler, 2010), as suggested by the Mplus developers (see http://www.statmodel.com/chidiff.shtml).

Figure 2.2. Conditional Piecewise Second-order Latent Growth Model of Student Social-Emotional Competence (SEC) Measured by Three DESSA-Mini Forms Across Nine Times of Measurement



Results

Descriptive Statistics

The average of eight DESSA-Mini items comprising each form increased, on average, by approximately 0.2 raw score points from Year 1 Fall to Year 3 Spring, on a 5-point Likert scale from 0=Never to 4=Very Frequently. When transformed into T-Scores based on the national norm (M = 50, SD = 10), this translates into about a 4 T-Score point increase throughout the three years. In Year 1 and Year 2, the increase in mean SEC scores occurred primarily between Fall and Winter of each year, whereas in Year 3, more increase occurred between Winter and Spring. There was a slight decrease in mean SEC scores during the Year 1 summer period (by 2.7 T-Score points) and less of such a decrease during the Year 2 summer (by 1.6 T-Score points). The distribution of the SEC scores was slightly left-skewed and platykurtic across all measurement time points. But no serious violations to univariate normality were identified as the absolute values of skewness and kurtosis were smaller than 1 for all items, which meets the customary thresholds for skewness and kurtosis (3 and 8, respectively; Kline, 2015; See Table 2.2).

Table 2.2. Descriptive Statistics of DESSA-Mini Ratings

Time	Valid N	Average of 8 items				T-Scores		Cronbach's α
		M	SD	Skewness	Kurtosis	M	SD	
Year 1 Fall	3,697	2.69	0.96	-0.43	-0.28	50.16	11.14	0.95
Year 1 Winter	3,149	2.85	0.95	-0.50	-0.33	53.18	11.72	0.96
Year 1 Spring	3,476	2.86	1.00	-0.56	-0.40	54.01	12.05	0.97
Year 2 Fall	3,744	2.76	0.97	-0.44	-0.39	51.29	11.75	0.96
Year 2 Winter	2,829	2.87	0.96	-0.47	-0.48	53.52	12.13	0.97
Year 2 Spring	3,656	2.84	1.02	-0.52	-0.55	53.87	12.49	0.98
Year 3 Fall	3,493	2.84	0.94	-0.49	-0.29	52.29	11.69	0.96
Year 3 Winter	2,263	2.82	0.98	-0.47	-0.45	52.91	12.24	0.97
Year 3 Spring	2,640	2.90	0.97	-0.58	-0.30	54.42	12.04	0.98

Note. The descriptive statistics were presented at the composite score level due to the space limitations. There was negligible item-level variation in these statistics. The number of valid cases did not differ at the item level at each occasion.

There was a substantial amount of missing data across time. Missing data pattern analysis revealed that 686 students (12.58%) had complete data for all the nine occasions, 784 (14.38%) were missing for one of the nine occasions, and 842 (15.44%) were missing for two of the nine occasions. About 9% of the students had some valid data for Year 1, but had missing data for all ratings in Year 2 and Year 3. About 14% of the students had some valid data for Year 1 and Year 2, but had missing data for Year 3. Given the high mobility rate of the district, it is likely that many of these cases represent students transferring out of the district during the study years. On the other hand, about 16% of the students had missing data for Year 1, but had some valid data for Year 2 and Year 3. About 7.56% of the students had missing data for all ratings in Year 1 and Year 2, but had some valid data in Year 3. Given the high mobility rate of the district, it is likely that many of these cases represent students transferring into the district during the study years. Table 2.3 presents the distribution of missing data patterns.

Additional tests revealed that students who had complete data tended to have higher SEC at baseline than those who had some missing data for later ratings ($t = 4.22$, $p < .001$), but the difference was negligible in size (Hedges' $g = .18$). The proportions of students having any missing data did not differ significantly by student gender nor SES, but differed by student race and ethnicity: compared to those who had complete data, there were slightly fewer Hispanic (by 3% point) or White students (by 2% point) and slightly more Black (by 3% point) or multi-racial or other race/ethnicity students (by 2% point) among those who had missing data ($\chi^2(3) = 8.80$, $p = .032$), but the difference was negligible in size (Cramer's $V = 0.04$). This study applied the FIML approach to missing data, when estimating the rate of growth while accounting for the baseline level differences and including race and ethnicity as one of the major predictors.

Table 2.3. Missing Data Patterns

Missing Pattern	Frequency	Percent
No missing across all three years	686	12.58
Missing earlier ratings		
Missing data for Year 1 + some valid data for Year 2 and Year 3	852	15.63
Missing data for Year 1 and Year 2 + some valid data for Year 3	412	7.56
Missing later ratings		
Some valid data for Year 1 + missing data for Year 1 and Year 2	512	9.39
Some valid data for Year 1 and Year 2 + missing data for Year 1	758	13.90
Missing inconsistently		
Missing at one time point across three years	784	14.38
Missing at two time points across three years	842	15.44
Other inconsistent missing patterns	606	11.12
Missing all ratings across three years	0	0.00
Total	5,452	100.00

Factorial Invariance Across Subgroups Testing Results

To examine the factorial invariance of the DESSA-Mini by student gender, race and ethnicity, and SES, a series of model fit comparisons with increasing invariance assumptions were conducted. Across grouping variables, although the chi-square difference tests were statistically significant, inappreciable changes were observed in practical fit indices between the configural invariance model and the weak factorial invariance model ($|\Delta CFI| \leq .00$, $|\Delta TLI| \leq .00$, $|\Delta RMSEA| \leq .00$), between the weak factorial invariance model and the strong factorial invariance model ($|\Delta CFI| \leq .00$, $|\Delta TLI| \leq .00$, $|\Delta RMSEA| \leq .00$), and between the strong factorial invariance model and the strict factorial invariance model ($|\Delta CFI| \leq .01$, $|\Delta TLI| \leq .00$, $|\Delta RMSEA| \leq .01$). These findings suggest that strict factorial invariance (i.e., metric, scalar, and residual invariance) can be assumed across gender, racial-ethnic, and socioeconomic subgroups, indicating that the entire factor model including factor loadings, intercepts, and residual variances was equal across subgroups. The strict factorial invariance model showed a good fit with each DESSA-Mini Form (CFI ≥ 0.93, TLI ≥ .93, RMSEA ≤ .04). The full invariance testing results are presented in Table 2.4.

Table 2.4. Fit Indices of a Series of Models With Increasing Invariance

Model [Comparison Model]	$\chi^2(df)$ [$\Delta\chi^2(\Delta df)$][a]	CFI [ΔCFI]	TLI [ΔTLI]	RMSEA [ΔRMSEA]
By Gender (female vs. male)				
Model 1: Configural invariance	18889.95(4494)***	0.94	0.93	0.03
[—]	[—]	[—]	[—]	[—]
Model 2: Weak factorial invariance	18987.93(4518)***	0.94	0.93	0.03
[vs. Model 1]	[97.56(24)***]	[.00]	[.00]	[.00]
Model 3: Strong factorial invariance	19509.47(4542)***	0.94	0.93	0.03
[vs. Model 2]	[574.77(24)***]	[.00]	[.00]	[.00]
Model 4: Strict factorial invariance	20308.60(4614)***	0.93	0.93	0.04
[vs. Model 3]	[609.39(72)***]	[-.01]	[.00]	[.01]
By Race/Ethnicity (Hispanic vs. non-Hispanic Black vs. non-Hispanic White)				
Model 1: Configural invariance	24330.89(6732)***	0.94	0.93	0.04
[—]	[—]	[—]	[—]	[—]
Model 2: Weak factorial invariance	24375.75(6780)***	0.94	0.93	0.04
[vs. Model 1]	[44.86(48)]	[.00]	[.00]	[.00]
Model 3: Strong factorial invariance	24629.55(6828)***	0.94	0.93	0.04
[vs. Model 2]	[253.80(48)***]	[.00]	[.00]	[.00]
Model 4: Strict factorial invariance	25099.37(6872)***	0.94	0.93	0.04
[vs. Model 3]	[469.82(144)***]	[.00]	[.00]	[.00]
By Socioeconomic Status (low-SES vs. middle-to-high-SES)				
Model 1: Configural invariance	19356.52(4494)***	0.94	0.93	0.04
[—]	[—]	[—]	[—]	[—]
Model 2: Weak factorial invariance	19417.25(4518)***	0.94	0.93	0.04
[vs. Model 1]	[52.51(24)**]	[.00]	[.00]	[.00]
Model 3: Strong factorial invariance	19563.06(4542)***	0.94	0.93	0.04
[vs. Model 2]	[152.19(24)***]	[.00]	[.00]	[.00]
Model 4: Strict factorial invariance	19552.63(4614)***	0.94	0.93	0.03
[vs. Model 3]	[225.09(72)**]	[.00]	[.00]	[-.00]

[a] The chi-square difference test statistics used the Satorra-Bentler scaling correction.

Note. Strong factorial invariance over time was assumed for each model.

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

SEC Growth Trajectory Modeling

After confirming factorial invariance across subgroups, the unconditional second-order piecewise latent growth model was first fitted to estimate the average SEC growth trajectory with three distinct growth curves for each year. Consistent with Lee et al. (2022b), the model that freely estimated the slope coefficients at the middle of each year, while assuming a decrease over the summer by a half of the previous year's gain, showed a good fit with the data: $\chi^2(2270) = 17228.91$, $p < .001$, CFI = 0.94, TLI = 0.93, RMSEA = 0.035, 90% CI [0.034, 0.035]. This model seemed to produce most reliable estimates with comparably good fit indices, compared to a few other competing models posing different assumptions about the summer loss (e.g., models assuming no decrease, different rates of decreases, decreases only for the first summer, etc.). The full specification and estimation results are presented in Table 2.5.

Table 2.5. Model Specification and Estimation Results of the Unconditional Piecewise Growth Model

	Estimate	SE	*p*	Estimate	SE	*p*	Estimate	SE	*p*
Coefficients of the Slope factor									
		$Slope_{Y1}$			$Slope_{Y2}$			$Slope_{Y3}$	
SEC_{Y1Fall}	0.00	—	—	0.00	—	—	0.00	—	—
$SEC_{Y1Winter}$	1.04	0.17	<.001	0.00	—	—	0.00	—	—
$SEC_{Y1Spring}$	1.00	—	—	0.00	—	—	0.00	—	—
SEC_{Y2Fall}	0.50	—	—	0.00	—	—	0.00	—	—
$SEC_{Y2Winter}$	0.50	—	—	1.07	0.12	<.001	0.00	—	—
$SEC_{Y2Spring}$	0.50	—	—	1.00	—	—	0.00	—	—
SEC_{Y3Fall}	0.50	—	—	0.50	—	—	0.00	—	—
$SEC_{Y3Winter}$	0.50	—	—	0.50	—	—	0.58	0.16	<.001
$SEC_{Y3Spring}$	0.50	—	—	0.50	—	—	1.00	—	—
Growth Factor Parameter Estimates									
		Means			Variances				
Intercept	0.00	—	—	0.56	0.03	<.001			
$Slope_{Y1}$	0.23	0.03	<.001	0.10	0.00	<.001			
$Slope_{Y2}$	0.14	0.02	<.001	0.24	0.05	<.001			
$Slope_{Y3}$	0.12	0.02	<.001	0.50	0.14	<.001			
Covariances									
		w/ $Slope_{Y1}$			w/ $Slope_{Y2}$			w/ $Slope_{Y3}$	
Intercept	0.01	0.02	0.66	-0.00	0.03	0.98	-0.08	0.02	<.001
$Slope_{Y1}$				-0.07	0.03	0.04	-0.01	0.02	0.81
$Slope_{Y2}$							-0.02	0.02	0.31

Table 2.5 (continued).

Residual Variances of SEC Latent Variables				*Expected Values of SEC Latent Variables*			
SEC_{Y1Fall}	0.46	0.03	<.001	SEC_{Y1Fall}	0.00	—	—
$SEC_{Y1Winter}$	0.37	0.03	<.001	$SEC_{Y1Winter}$	0.24	0.04	<.001
$SEC_{Y1Spring}$	0.28	0.02	<.001	$SEC_{Y1Spring}$	0.23	0.03	<.001
SEC_{Y2Fall}	0.55	0.02	<.001	SEC_{Y2Fall}	0.12	0.01	<.001
$SEC_{Y2Winter}$	0.30	0.03	<.001	$SEC_{Y2Winter}$	0.26	0.03	<.001
$SEC_{Y2Spring}$	0.27	0.02	<.001	$SEC_{Y2Spring}$	0.25	0.03	<.001
SEC_{Y3Fall}	0.52	0.02	<.001	SEC_{Y3Fall}	0.18	0.02	<.001
$SEC_{Y3Winter}$	0.37	0.04	<.001	$SEC_{Y3Winter}$	0.25	0.04	<.001
$SEC_{Y3Spring}$	0.06	0.13	.660	$SEC_{Y3Spring}$	0.30	0.03	<.001

Note. The coefficients for the intercept factor were all fixed at 1. The mean of the intercept factor is fixed at 0, and the mean of each slope factor indicates the mean change scores between Fall and Spring within the corresponding year relative to the standard deviation of SEC scores at baseline. The results for the first-order components of the model are not presented here for simplicity.

To examine the extent to which student SEC growth trajectories differ across student subgroups, gender, race and ethnicity, and SES were included as a set of predictors of the growth model examined above, along with student grade level and the PATHS implementation. This conditional second-order piecewise latent growth model showed a good fit: $\chi^2(2678) = 17687.19$, $p < .001$, CFI = 0.935, TLI = 0.927, RMSEA = 0.033, 90% CI [0.033, 0.034]. Table 2.6 presents the regression coefficients on the intercept and three slope factors.

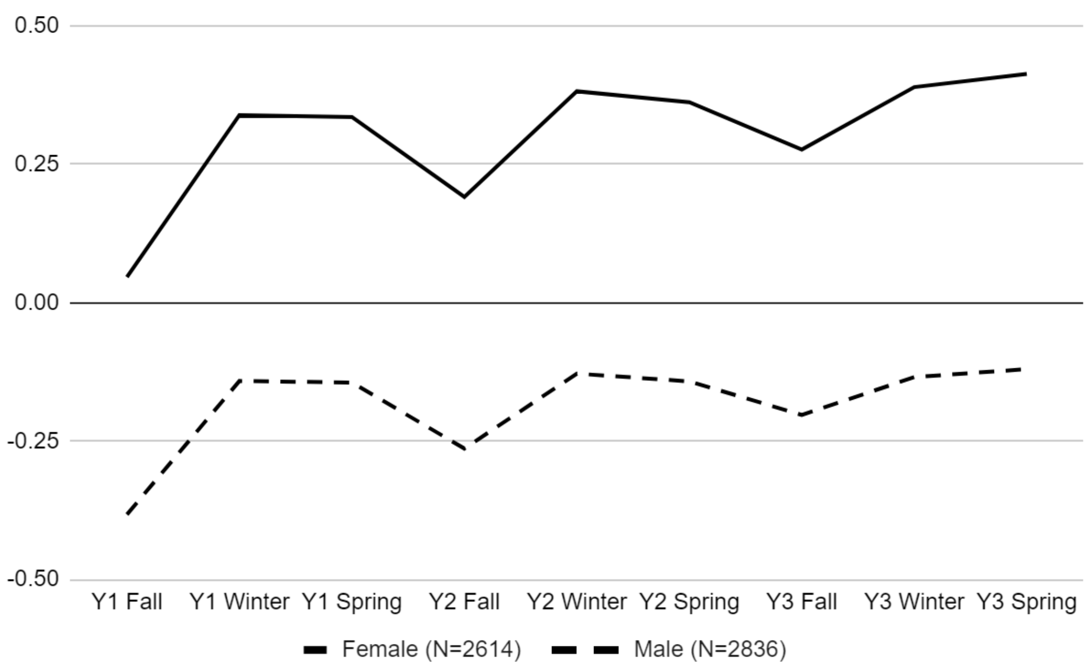
Table 2.6. Estimated Regression Coefficients of the Conditional Piecewise Growth Model

	Est.	SE	*p*	Est.	SE	*p*	Est.	SE	*p*
		Intercept (R^2=.14)							
Male	-0.43	0.03	<.001						
Black (vs. White)	-0.17	0.05	.002						
Hispanic (vs. White)	-0.14	0.04	.001						
Low-income	-0.15	0.05	.001						
Grade	0.18	0.02	<.001						
PATHS-continued	0.16	0.04	<.001						
		$Slope_{Y1}$ (R^2=.19)			$Slope_{Y2}$ (R^2=.06)			$Slope_{Y3}$ (R^2=.03)	
Male	-0.05	0.03	.085	-0.05	0.03	.050	-0.05	0.03	.075
Black (vs. White)	-0.11	0.05	.028	0.03	0.05	.523	-0.04	0.06	.488
Hispanic (vs. White)	0.01	0.04	.888	0.05	0.04	.165	0.06	0.05	.215
Low-SES	0.05	0.04	.221	-0.01	0.04	.821	-0.05	0.05	.215
Grade	-0.17	0.02	<.001	-0.10	0.02	<.001	-0.02	0.02	.239
PATHS-continued	0.00	0.04	.958	0.16	0.03	<.001	0.17	0.04	<.001

At baseline, significant gender, racial-ethnic, and socioeconomic differences in SEC were found. Male (vs. female; b = -0.43, $p < .001$), Black (vs. White; b = -0.17, $p = .002$) or Hispanic (vs. White; b = -0.14, $p = .001$), and low-income (vs. middle-to-high income; b = -0.15, $p = .001$) students each had lower mean intercepts in their SEC growth trajectories. Male students had marginally lower mean rates of growth than female students throughout the three years (Year 1: b = -0.05, p < .10; Year 2: b = -0.05, p = .50; Year 3: b = -0.05, p < .10). A difference in the mean growth rate by race and ethnicity was only found between Black and White students in Year 1. Black students showed a lower mean rate of growth than White students (b = -0.11, p < .05) in Year 1. Student SES was not significantly associated with the growth rates throughout the study years. Figures 2.3 to 2.5 show the estimated marginal means of the SEC factor over time across gender, racial-ethnic, and socioeconomic subgroups.

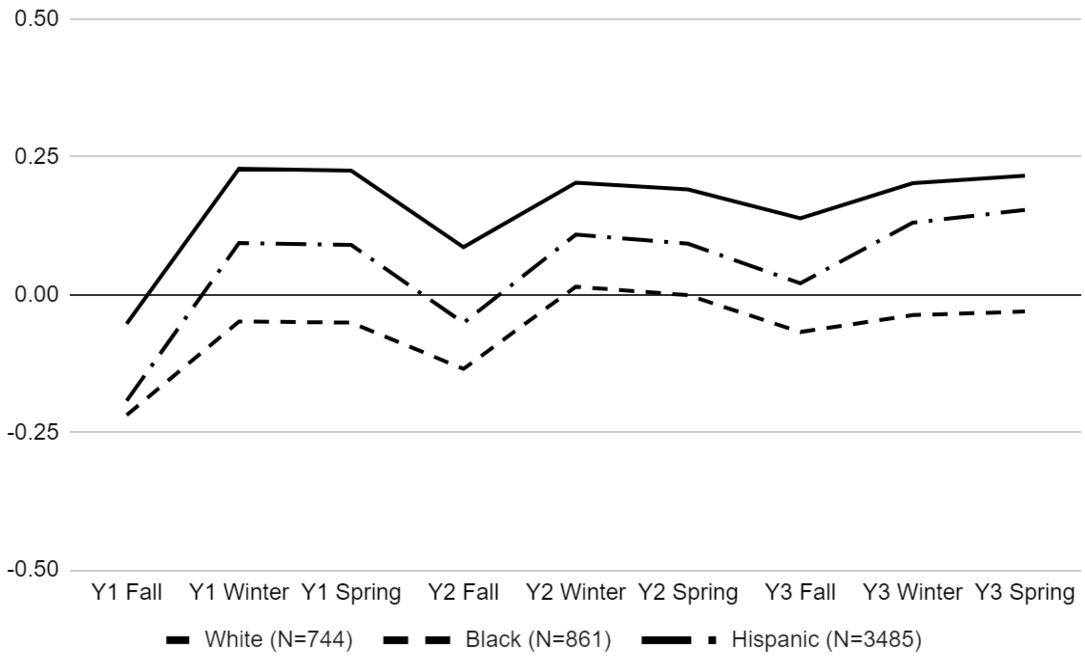
Student grade level and the exposure to continued PATHS implementation were also related to the student SEC growth trajectory. Students in higher grades had a higher mean baseline (b = 0.18, p < .001), but their mean rates of growth in the first two years were slower than lower-grade students (Year 1: b = -0.17, p < .001; Year 2: b = -0.10, p < .001). The Year 3 mean growth rate did not differ by grade levels. Students who were exposed to continued PATHS implementation for three years (vs. discontinued in Year 2 and Year 3) showed a larger mean growth in Year 2 (b = 0.16, p < .001) and Year 3 (b = 0.17, p < .001), while they had a higher mean baseline (b = 0.16, p < .001) and similar mean rate of growth in Year 1. Altogether, this conditional piecewise latent growth model explained 14% of the baseline SEC level, 19% of the Year 1 growth rate, 6% of the Year 2 growth rate, and 3% of the Year 3 growth rate.

Figure 2.3. Marginal Means of SEC Over Time Across Gender Subgroups



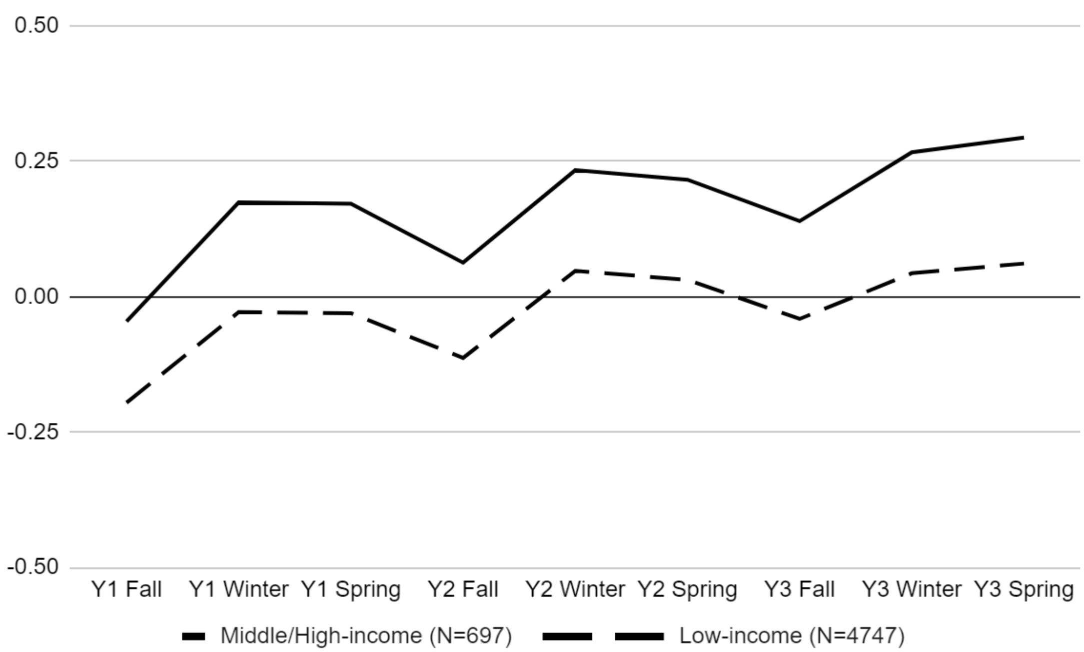
Note: Y-axis indicates the relative group difference in standard deviation units.

Figure 2.4. Marginal Means of SEC Over Time Across Racial-Ethnic Subgroups



Note: Y-axis indicates the relative group difference in standard deviation units.

Figure 2.5. Marginal Means of SEC Over Time Across Socioeconomic Subgroups



Note: Y-axis indicates the relative group difference in standard deviation units.

Discussion

The goal of this study was to contribute to advancing research on disparities in student mental health by measuring and identifying gender, racial-ethnic, and socioeconomic disparities in social-emotional competence (SEC) developmental trajectories during elementary school years. Specifically, this study first tested measurement invariance of a widely-used teacher-completed behavioral rating scale for SEC assessment across student gender, race and ethnicity, and socioeconomic status (SES), and then examined the extent to which students' multiyear SEC growth trajectories differ across these subgroups under a routine, district-wide SEL implementation condition. The invariance testing results supported strict factorial invariance of the DESSA-Mini across all the examined subgroups throughout nine occasions during three elementary school years, providing a foundation for valid cross-group comparisons of student SEC growth trajectories. The growth modeling results indicated that male (vs. female), Black or Hispanic (vs. White), and lower-income (vs. higher-income) students started with a lower level of SEC, and these gaps were either sustained across time (between Hispanic and White, and lower-income and higher-income students) or slightly widened (between males and females throughout all three years, and between Black and White students during the first year). Taken together, these findings provide robust evidence of differences in the development of SEC, one of the critical protective and promotive factors of student mental health among other important outcomes. Implications for research and practice as well as limitations of this study are discussed next.

Measurement invariance across subgroups is an essential but often overlooked criteria for studying subgroup health disparities (Meredith & Teresi, 2006). This gap is also evident in the literature on social and emotional learning (SEL). Although teacher-completed behavioral rating scales are now widely used as universal screening and progress monitoring tools in the domain of SEL, there is a lack of studies that test measurement invariance. This is the first study that explicitly tested the factorial invariance of the DESSA-Mini across student subgroups and, to my knowledge, any teacher-completed behavioral rating scales currently used for universal SEC assessment in educational practice (e.g., as identified in the SEL Assessment Guide; Assessment Work Group, 2021). The evidence supporting strict invariance suggests that teachers rate all the DESSA-Mini items in the same way for students who have the same level of SEC, regardless of student gender, race and ethnicity, and SES. Therefore, the observed subgroups differences in the DESSA-Mini scores can be interpreted as reflecting real differences in the construct measured, rather than teachers' differential rating behaviors based on student identities.

This finding not only allows researchers to make valid cross-group comparisons to further investigate subgroup disparities in SEC, but also contributes to reducing the field's widespread concern about the potential of measurement bias when assessing diverse student bodies (Shapiro et al., 2016). Given ample evidence of teacher bias in reporting "problem" behaviors resulting in discriminatory disciplinary practices against marginalized students, particularly based on their race and ethnicity (Bradshaw et al., 2010; Gregory & Roberts, 2017; Skiba et al., 2002, 2011; Ura & d'Abreu, 2022), the lack of measurement bias found in this study may be surprising. It should be noted, however, that this study only tested one aspect of measurement bias, measurement invariance, and only of one particular behavioral rating scale, composed exclusively of strength-based items (i.e., indicating positive behavior rather than maladaptive behavior). In fact, a review of 13 studies that examined teacher-completed behavioral rating scales, compared against a third-party criterion measures of behavior, suggested that there is a lack of consistent evidence supporting the presence of racial-ethnic bias in teacher ratings of student internalizing or externalizing behavior (Mason et al., 2014), despite the disproportionality of referral. Taken together, this study calls for more explicit testing of measurement invariance of various behavioral rating scales for SEC assessment, especially the ones that are actively used in practice. This evidence should guide the selection of assessment tools to be implemented and be transparently made available to all stakeholders and rights-holders. With more data available, factors associated with measurement invariance vs. non-invariance could be further explored across different tools, which will also inform the development and improvement of SEC assessments that are less prone to teacher bias. Also, if measurement invariance findings are replicated for SEC assessments, the field may benefit from a greater focus on score *interpretation* and *fair utilization* to find paths to reduce disproportionate representation in stigmatizing, exclusionary, or punitive practices.

Nevertheless, the evidence of measurement invariance does not guarantee that the measure is free of bias and equally relevant to all student subgroups. For example, the Assessment Work Group (2019) has raised an issue of construct underrepresentation, which occurs when certain manifestations of SEC that are critical to understand specific subgroups' SEC are left out in the assessment (e.g., racial-ethnic identity and code-switching skills for students of color; Duchesneau, 2020). In this case, even if teachers rate the included items in the same way across subgroups, the rating scores may not capture the full spectrum of SEC for certain subgroups. A recent article has proposed the potential utility of prototype analysis that combines both qualitative content validation and quantitative invariance testing in order to develop SEC measures that represent diverse cultural assets (El Mallah, 2020). Use of a multi-informant approach is also recommended to have a more comprehensive understanding of student SEC. This study calls for more research that contributes to the evidence base for equitable SEC assessments; Measurement invariance testing of a behavioral rating scale is just one of the necessary steps that has been under-investigated, and much more work is needed to make sure that SEC assessment tools are equally relevant and fairly applied across diverse subgroups of students.

This study also contributes to the knowledge base on disparities in SEC developmental trajectories during school years. The observed gender, racial-ethnic, and socioeconomic disparities at baseline are generally consistent with previous findings in available literature (e.g., Aber et al., 2003; S. M. Jones et al., 2011; West et al., 2020), with added confidence that they are not an artifact of teachers' differential ratings. This study also provides fairly robust evidence that these disparities were not reduced throughout multiple years, even in the context of a districtwide SEL initiative. The effect sizes for baseline subgroup disparities were small for gender (ES = .43, $p < .001$) and negligible for race, ethnicity, and SES (ES = [.14, .17]) per Cohen's (1969) criteria. However, the use of Cohen's standards has been criticized to be too stringent for educational data (Kraft, 2019). Given that the estimated effect sizes for yearly

SEC growth in this study ranged from .12 to .17, and the total average increase was estimated to be .30, it is quite alarming that male students, on average, lagged behind female students by more than three years of difference and that students of color and low-income students, on average, lagged behind their White and higher-income peers by about one year of difference.

Among the three goals of research on disparities—detecting, understanding, and reducing disparities, the focus of the present study is on disparity detection. It is beyond the scope of this study to explain the causes and mechanisms underlying these subgroup disparities. However, it should be emphasized that there is no reason to believe that there are any inherent differences in SEC by child gender, race and ethnicity, and SES. Rather, the subgroup disparities observed in this study may be understood as a reflection of different socialization and/or marginalization experiences based on these characteristics. The evidence of subgroup disparities in SEC must not be misused to reinforce existing prejudices and systems of oppression. Just as the academic achievement gap should be reframed as the educational debts that the society owes its students (Ladson-Billings, 2006), I hope this observation of SEC disparities can inform practice decisions and policy-making efforts to effectively address the underlying social inequalities and ensure more equitable social-emotional developmental environments and outcomes.

From this standpoint, the findings of the current study raise many interesting questions as to why male, Black or Hispanic, or low-income students, on average, consistently showed lower levels of SEC than their counterparts during elementary school years. Regarding gender gap, differential norms and expectations for social-emotional behavior based on gender (e.g., "boys will be boys") and social pressure to conform to gender stereotypes (e.g., "boys are stronger both physically and emotionally") might be responsible for the observed gender disparities in SEC (Skipper & Fox, 2021). In fact, some studies with younger children have found that parents' and preschool teachers' emotion socialization behaviors were differential based on child gender, leading to girls generally having higher emotional competence than boys (Denham et al., 2010; King, 2021). Gender stereotyping may not just hinder male students' positive social-emotional development, but also can take a toll on female students in the longer term. For example, in a large-scale self-report survey study that simulated social-emotional growth trajectories of students in Grade 4 to 12 (West et al., 2020), female students were found to have consistently higher levels of self-management and social awareness than male students, but their self-efficacy started to become lower than males from Grade 6, and this gap increasingly widened over time. More research is called for to better understand the mechanisms underlying these gender disparities in SEC development and the implications for promoting gender equity in mental and behavioral health.

Although smaller in size than gender disparities, racial-ethnic and socioeconomic disparities observed in this study require special attention as structurally disadvantaged subgroups were found to have lower levels of SEC than structurally advantaged subgroups. These gaps might be an indication of the society's racial, ethnic, and economic inequities that shape unequal developmental conditions. Literature has suggested that students of color and low-income students tend to be more frequently exposed to stressful events and conditions which induce toxic stress that impedes children's healthy cognitive, emotional, and behavioral development because of, for example, racial-ethnic discrimination and poverty (Morsy & Rothstein, 2019). Also, particularly with regard to race and ethnicity, cultural misalignment or dissonance between home and school has been suggested as an important factor to consider when understanding seemingly less adaptive (more precisely, less conforming) social-emotional behavior among students of color within a school culture endorsing white, anglocentric values (Arunkumar et al., 1999; Boykin et al., 2005). This also ties back to the question of bias in SEL assessments—whether culturally diverse manifestations of SEC are equally promoted and fully assessed in the current practices of SEL. Again, further research is warranted to understand the underlying causes and mechanisms behind these racial, ethnic, and socioeconomic disparities in order to create more equitable school environments and challenge social inequities at large.

This study has several limitations to note. First, as in other studies using school administrative data to measure student identities, the simplistic categorization of gender, race and ethnicity, and SES does not respect students' own views of their identities and obscures within-group variability. Second, this study cannot answer how intersectional identities based on gender, race and ethnicity, SES, and many other constructs are linked to student SEC development. In fact, the challenges of defining and measuring subgroups (e.g., Perez & Hirschman, 2009) as well as examining intersectionality (Bauer, 2014; Bowleg, 2012) in quantitative research is well documented in the broad literature. Also, a substantial amount of variability in student SEC growth trajectories remained unexplained by the examined model. This study generates questions about how to advance quantitative research on subgroup disparities to better represent real-world complexities. Based on the evidence indicating the presence of inequalities in SEC developmental trajectories, I call for more nuanced and sophisticated research that can further advance our understanding of subgroup disparities in SEC development and lead to appropriate societal interventions.

Another important limitation to note is that, with the observational study design, it was not possible to estimate counterfactual growth trajectories under the condition where no SEL initiative was implemented. Therefore, the fact that the disparities were not reduced across school years and even widened for some subgroups does not reveal anything about the effects of SEL practices or development under non-intervention conditions. In order to understand whether an SEL intervention reduces, sustains, or increases existing disparities, experimental or quasi-experimental studies involving a reasonable comparison condition are required.

Among other factors regarding limited generalizability (e.g., time, location), underrepresentation of other racial-ethnic groups should also be noted. In this study that analyzed secondary data, SEC growth trajectories of students identified as American Indian or Alaska Native, Asian, Hawaiian or Pacific Islander, and multi-racial could not be reliably estimated due to small group sizes. However, small group size should not be a perpetual excuse for not investigating minority (or minoritized) groups' outcomes. More research representing diverse student populations is needed to advance epistemic justice.

Despite all these limitations, this study provides fairly robust evidence of existing subgroup disparities in SEC developmental trajectories, using three years of longitudinal data and advanced statistical methods. It also provides the groundwork for an agenda for future research and practices in the field of SEL, such as how to determine whether SEL assessments are equitable and fair across diverse student bodies, how to understand the observed subgroup disparities in SEC development, and how to effectively tackle these inequities – not by blaming individual students or by trying a "quick fix" (Duchesneau, 2020), but through addressing the root causes of these inequities in the school system and beyond.

To address one of the limitations of this study, Chapter 3 examines the universality of the intervention effects of universal SEL across diverse student subgroups, and the effects of universal SEL on existing subgroup disparities.

Chapter 3

Universal school-based social and emotional learning (SEL) for diverse student subgroups: Implications for enhancing equity through SEL

Summary

School-based, universal social and emotional learning (SEL) has been widely practiced and promoted as a promising approach to prevent youth mental, emotional, and behavioral problems. Although prior research has accumulated robust evidence of the average intervention effects of universal SEL, it remains unclear whether universal SEL works similarly or differentially across diverse sociocultural subgroups of students. This has implications for understanding the impact of universal SEL on possible subgroup disparities in student social-emotional competence (SEC). This chapter presents a study that examines whether the effects of universal SEL on student SEC growth differed across diverse student subgroups classified by gender, race, ethnicity, socioeconomic status, disability status, and English learner status. Data came from student SEC progress monitoring collected during a one-year quasi-experimental study of a universal SEL program (N = 1,592; Grades K-2). The results of multigroup latent growth modeling suggest that (a) the intervention effects were slightly larger for Black students, compared to White or other racial-ethnic subgroups, and (b) the effects were not different across other examined subgroups. This study also found that in the comparison condition, the SEC disparities between Black and White students tended to widen throughout the year, whereas in the intervention condition, Black students showed a similar rate of growth as their White peers. Findings suggest that universal SEL may prevent racial disparities from further widening, but it may not be sufficient to reduce existing subgroup disparities. More studies should examine differential effects of universal preventive programs by diverse subgroups to better inform practices that enhance equity in youth outcomes.

Background

Mental, emotional, and behavioral (MEB) well-being is important for children and adolescents to thrive and to become healthy, prosocial adults. Despite demonstrations that prevention of MEB health problems is possible, still too many young people experience problems that threaten their well-being (National Academies of Sciences, Engineering, and Medicine, 2019). In the United States and many other countries, school-based social and emotional learning (SEL) has been widely adopted as a promising approach to universal prevention of a wide range of youth MEB health problems. This is predicated on the literature suggesting that social-emotional competence (SEC), the primary intervention target of SEL, is a common and malleable protective factor for healthy youth development (Domitrovich et al., 2017). Several meta-analyses have provided evidence supporting the overall benefits of universal SEL in promoting greater mean levels of student SEC and other positive developmental outcomes (e.g., Durlak et al., 2011; Taylor et al., 2017). Based on this, systemic efforts have been made to scale up SEL at the district, state, and national levels (Greenberg et al., 2017).

Potential Effects of Universal SEL on Disparities in Social-Emotional Competence

Less is known about whether universal SEL programming works equally well across diverse sociocultural subgroups of students. Subgroups, in this use, are classifications of people into groups (e.g., gender, race) based on socioculturally constructed identities, grouped by the attribution or actuality of a set of shared characteristics or experiences. In the broader literature of prevention science, the need for testing the *universality* of intervention effects (i.e., effectiveness for all people) has been well articulated (Perrino et al., 2015). For example, the Standards of Evidence for Efficacy, Effectiveness, and Scale-up endorsed by the Society of Prevention Research clearly state that the demonstration of intervention effects across population subgroups is a prerequisite for scaling up evidence-based interventions (Gottfredson et al., 2015). Although subgroup analyses are no longer exceedingly rare (e.g., Kim et al., 2015), the analysis of subgroup effects for most distinguished universal preventive programs remains lacking (Shapiro et al., 2022).

A universal intervention that improves protective factors, on average, does not *necessarily* reduce disparities (i.e., status differences) in protective factors between subgroups or enhance equity–the other important goals of prevention in addition to improving the average outcome. To illustrate some possible effects of universal SEL on disparities in SEC, let's assume that we initially observe disparities in SEC between a historically marginalized and privileged subgroup, and without any SEL intervention, these disparities are sustained over time. Conceptually, adding an effective-on-average universal SEL intervention could have at least three impacts on SEC disparities, depending on how its benefits are distributed across subgroups.

Three illustrative impacts are as follows. First, if the intervention brought some benefits to all subgroups, but the amount of benefits was larger for marginalized groups, the intervention would *improve* average outcome and *decrease* disparities in the outcome (Alvidrez et al., 2019). Second, if the intervention was equally beneficial, or brought the same amount of positive effects to students across subgroups, the intervention would *improve* average outcomes and *maintain* the SEC disparities observed at baseline (Shapiro, Derr, et al., 2022). One might argue that although the intervention did not achieve the goal of reducing disparities, it is still worthy of dissemination because of its effects on promoting population-level outcomes, with positive effects for all subgroups of students. Third, if the intervention benefited all subgroups, but the benefits were larger for privileged subgroups, the intervention would improve average outcomes and *exacerbate* the outcome disparities. This approach produces "intervention-generated inequalities" (Lorenc et al., 2013, p. 190). The notion of Inequality Paradox, proposed by Frohlich & Potvin (2008), suggests that universal preventive interventions may inadvertently increase health disparities by bringing larger benefits to those at lower risk. In this case, despite its demonstrated effects on improving the average outcome, scaling up an intervention that increases subgroup disparities may be unjustifiable from an equity-oriented perspective. These simplistic examples only illustrate scenarios in which all groups benefit (albeit to varying extents), and ignore that program selection has many considerations (e.g., how much a group benefits from an intervention *relative* to other available options); Real world complexities only enhance the argument for transparent reporting of subgroup differences to maximize information for participatory decision making.

Empirical Evidence on SEL Intervention Effects Across Sociocultural Subgroups

The SEL literature first focused on evaluating average program effects on the general student body in order to understand *what works*. Studies then attempted to understand *for whom* SEL works. To date, this has largely been done by comparing the relative sizes of intervention effects (1) across diverse contexts and samples (e.g., across study sites with different demographic compositions) and (2) across subgroups classified by different levels of social-emotional needs or risks. To the first point, meta-analyses of universal, school-based SEL have suggested that intervention effects may be similar across schools with different demographic compositions in terms of student gender, race, and ethnicity, and that SEL may yield more benefits to students in schools with a higher proportion of low-income students (Durlak et al., 2011; Taylor et al., 2017; Wilson & Lipsey, 2007). However, these meta-analytic findings do not address the extent to which universal SEL works equally across diverse subgroups of students *within* the same intervention condition. A positive average intervention effect found in a school does not explain how the benefits are distributed across subgroups of students within the school. Too often, by only reporting the aggregated average effect, a homogeneous effect across subgroups is assumed without confirmation.

To the second point, some SEL evaluation studies have examined the extent to which universal SEL provides relative benefits to student subgroups classified by different levels of risk. A plurality of these studies have found that students with higher social-emotional needs or elevated risk at baseline (e.g., higher level of aggression at pre-intervention) tended to benefit more from a universal SEL intervention (Bierman et al., 2010; Carroll et al., 2020; Frey, Nolen, et al., 2005; Low et al., 2019; Malti et al., 2011), while one study found that the benefits of a universal SEL program were similar across student subpopulations classified by different social-emotional and behavioral developmental trajectories (e.g., relatively higher vs. lower risk levels; Duncan et al., 2017). Although it can be difficult to know whether these findings are manifestations of their study designs or other confounding factors (e.g., regression to the mean, maturation), at face value these findings may suggest that universal SEL has the potential to reduce, rather than increase, the disparities in outcomes. Yet, disaggregating intervention effects by risk status does not determine whether SEL maintains, reduces, or exacerbates any existing SEC disparities across socioculturally classified subgroups.

Direct evidence as to whether SEL intervention effects are similar across diverse sociocultural subgroups has been limited and inconclusive. According to Rowe and Trickett's (2018) review of the 117 peer-reviewed articles that were included in Durlak et al.'s (2011) meta-analysis of universal school-based SEL, 37 studies conducted subgroup analysis by gender, and only 8 studies by race and ethnicity, 2 studies by socioeconomic status, and 3 studies by disability status, including both stratified analysis (i.e., separate tests of intervention effects for each subgroup) and moderation analysis involving significance testing (i.e., testing whether intervention effects significantly differ by subgroup). Among the relatively small number of studies that conducted subgroup analysis, mixed findings were observed regarding whether the intervention effect was moderated or not. Even when significant moderation effects were found, their directions were inconsistent across studies.

Rowe and Trickett (2018) deemed the overall paucity of student characteristics information, and the lack of appropriate subgroup analyses, in the reviewed studies "troubling" given the current scaling-up movements in the field of SEL (p. 573). They also advised "considerable caution about generalizing the overall positive findings of the meta-analysis to diverse groups of program participants" (p. 572). Without sufficient evidence suggesting that universal SEL programming may at least work equally well across diverse subgroups or may bring any benefits to marginalized subgroups, scaling up SEL may yield the unintended consequence of exacerbating inequalities.

Research Objectives

The main objective of this study is to examine whether an example universal SEL intervention works similarly or differentially across diverse subgroups of students, and discuss its implications for subgroup disparities in SEC. Specifically, this study examined the extent to which the one-year implementation of an SEL program, found to be effective on average, was differentially related to student SEC growth trajectories across diverse student subgroups under a quasi-experimental condition.

Based on a small body of conceptual and empirical work in the SEL literature suggesting diverse sociocultural characteristics that might be associated with different levels of SEC (e.g., Garner et al., 2014; Jones et al., 2020; Lee et al., 2022a; O'Brennan et al., 2015; Rowe & Trickett, 2018), this study examined student gender, race and ethnicity, socioeconomic status, disability status, and English language learner status as characteristics classifying student subgroups. This literature primarily documents gender disparities (where females have higher protection), is inadequate in the face of community concerns about racial and class disparities, and is largely silent on disparities related to disability status and English language learning (although social-emotional needs of students in special education and English language learners are well-described in the broader literature). Also, given the limited and mixed empirical evidence documenting the extent to which these disparities change under SEL intervention conditions, no specific hypothesis was set regarding differential intervention effects across these subgroups.

Method

Data and Sample

Data used in this study come from a quasi-experimental study of a school-based universal SEL program, TOOLBOX^{TM}, conducted within a single school district in the academic year of 2015-2016. TOOLBOX is a classroom-based universal SEL program (M. A. Collins, 2015), adopted in more than 40 school districts in Northern California and found to be effective in promoting K-2 students' SEC, on average, in routine school settings where no additional coaching or technical assistance was provided after a pre-intervention six-hour in-service training (Shapiro, Lee, et al., 2022). During the 2015-2016 academic year, four elementary schools were assigned to implement the TOOLBOX program, while two schools served as comparison (non-TOOLBOX) schools, all within the same school district in California. The assignment to conditions was not random but conducted in a way that intentionally distributed student demographic characteristics as evenly as possible across conditions. This district serves a racially and ethnically diverse and relatively low-income student body. During the study year, the district's total elementary school enrollment was 5,064, 59.1% of whom were identified as Hispanic, 13.0% Asian, 10.5% Black, and 8.0% White, with 71.3% receiving free or reduced-price lunch, and 39.8% identified as English language learners (District, 2016).

The current study focuses on K-2 students under both intervention (TOOLBOX) and comparison (non-TOOLBOX) conditions, for whom a positive intervention effect on average SEC has been demonstrated (Shapiro et al., 2022). The current study sample consists of 1,592 students, excluding 174 students from the original project sample belonging to extremely low prevalence racial/ethnic identities, such that measurement invariance across subgroups could not be tested (more detailed information is described in the following measurement section). In the current study sample, 65.95% were under the intervention condition and 34.05% were under the comparison condition. Table 3.1 shows the characteristics of the sample by intervention condition. Grade and gender were quite evenly distributed (35.7% Kindergartners, 31.2% first graders, and 33.0% second graders; 48.6% female and 51.4% male). Sixty three percent of the sample was identified as Hispanic, 15.5% as Asian, 12.9% as Black, and 8.6% as White. Seventy three percent of the sample was identified as low-income, indicated by their eligibility for free or reduced-price lunch. One student who had missing information for free or reduced-price lunch eligibility was excluded from analysis. Approximately 8% of the sample was receiving special education services (SPED; their disability code was primarily speech or language impairment [54.7%], autism [22.7%], and specific learning disability [11.7%]). About half of the sample (52.01%) was identified as English language learners (ELL; their primary language used at home was mostly Spanish [74.8%], followed by Cantonese [13.3%] and 17 additional languages). The distributions of grade, gender, SES, disability status, and ELL status did not differ significantly between conditions. The race and ethnicity distribution was found to be slightly different between conditions ($\chi^2(3) = 20.86$, $p < .001$, Cramer's V = .11 indicating a small difference). When comparing SEC growth trajectories between conditions, any differences associated with race and ethnicity (and all the other described student-level variables) were accounted for.

Table 3.1. Sample Characteristics by Intervention Condition

Variable	Category	Intervention (TOOLBOX) Sample		Comparison (non-TOOLBOX) Sample		Total	
		N	%	N	%	N	%
Grade	K	379	36.10	190	35.06	569	35.74
	1	320	30.48	177	32.66	497	31.22
	2	351	33.43	175	32.29	526	33.04
Gender	Female	504	48.00	270	49.82	774	48.62
	Male	546	52.00	272	50.18	818	51.38
Race and Ethnicity	Asian	176	16.76	70	12.92	246	15.45
	Black	158	15.05	47	8.67	205	12.88
	Hispanic	625	59.52	379	69.93	1,004	63.07
	White	91	8.67	46	8.49	137	8.61
Socio-economic Status (SES)	Middle-to-high income	294	28.00	131	24.17	425	26.70
	Low-income	755	71.90	411	75.83	1,166	73.24
	Missing	1	0.10	-	-	1	0.06
Disability Status	Not receiving SPED	966	92.00	498	91.88	1,464	91.96
	Receiving SPED	84	8.00	44	8.12	128	8.04
English language learner (ELL) Status	Non-ELL	526	50.10	238	43.91	764	47.99
	ELL	524	49.90	304	56.09	828	52.01
	Total	1050	100.00	542	100.00	1,592	100.00

Measurement

Student SEC was repeatedly measured using three forms of the DESSA-Mini (Naglieri et al., 2011/2014), three times a year (Form 1 in Fall, Form 2 in Winter, and Form 3 in Spring). Fall ratings were collected around October, Winter ratings around January, and Spring ratings around June. The DESSA-Mini is a brief (8-item) version of the full 72-item DESSA (LeBuffe et al., 2009/2014), a strength-based, standardized, and norm-referenced behavioral rating scale, designed to be used for universal screening and progress monitoring in the context of a large-scale implementation of SEL (LeBuffe et al., 2018). Each DESSA-Mini form comprises eight non-overlapping items asking teachers to rate the frequency of a student's behavior observed during the past four weeks on a 5-point Likert scale (0=Never to 4=Very frequently). Example items include "(how often did the child) keep trying when unsuccessful," "respect another person's opinion," "do something nice for somebody," and "show good judgment." The sum of 8 items is transformed into T-Scores based on a nationally representative standardization sample (M = 50, SD = 10). According to the DESSA-Mini manual, T-Scores of 60 and above indicate that the student has *Strengths*, T-Scores between 41 and 59 represent *Typical* SEC, and T-Scores of 40 and below indicate a *Need for Instruction*.

The DESSA-Mini has demonstrated strong psychometric qualities including reliability (e.g., internal reliability, test-retest reliability, inter-rater reliability, alternate form reliability) and validity (e.g., concurrent and predictive criterion validity) as summarized elsewhere (LeBuffe et al., 2018; Shapiro, Kim, et al., 2017). The DESSA-Mini has also demonstrated strong factorial invariance over time and across gender, racial-ethnic, and socioeconomic subgroups

(Lee et al., 2022a, 2022b). In the current study, the internal reliability coefficient estimates (Cronbach's alpha) ranged .95-.96 across three occasions. Using the current study sample, the three forms of the DESSA-Mini were tested to show strong factorial invariance across gender (female vs. male), race and ethnicity (Asian vs. Black vs. Hispanic vs. White), SES (low-income vs. middle-to-high-income), disability status (receiving SPED vs. not), and ELL status (ELL vs. non-ELL). Strong measurement invariance is a prerequisite for interpreting any observed subgroup differences in ratings as reflecting real subgroup disparities in the latent variable (SEC) rather than attributable to measurement bias. When testing measurement invariance by race and ethnicity, inclusion of smaller racial-ethnic subgroups other than Asian, Black, Hispanic, or White (either each as a unique subgroup or all together as a fifth group) yielded model convergence problems. Although the inclusion of these students in the main analyses led to similar conclusions about differential intervention effects across subgroups of interest, the findings presented in this chapter are based on the sample comprising only the four racial-ethnic subgroups for which evidence of measurement invariance has been established.

Student gender, race and ethnicity, SES, disability status, and ELL status were collected from an administrative database and dummy coded in ways that students had the value of 1 when they were male (vs. female), Asian, Black, or Hispanic (vs. White), low-income (vs. middle-to-high-income), receiving SPED (vs. not receiving), and ELL (vs. non-ELL), respectively.

Analysis Plan

This study conducted multigroup latent growth modeling, an application of multigroup structural equation modeling (SEM) to modeling growth trajectories over time. Multigroup (or multisample) SEM is used when the main question concerns whether the model parameters of interest differ across multiple groups, or in other words, whether group membership moderates the relations among variables specified in the model (Kline, 2015).

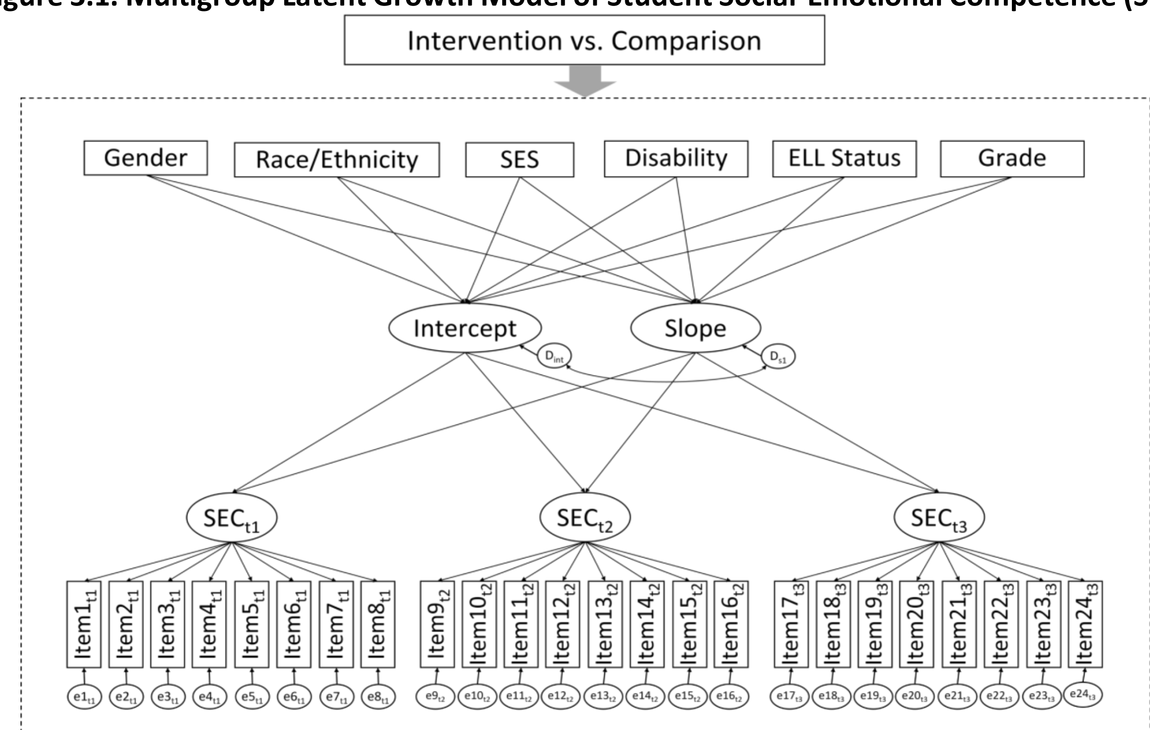
This study used two parallel modeling strategies with different parametrization–one parameterizing intervention effects and the other subgroup disparities in SEC growth trajectory. In the first approach, a set of multigroup analyses were conducted across sociocultural subgroups of interest to examine whether the effects of SEL intervention were similar or different across diverse subgroups of students. In this approach, a sociocultural characteristic of focal interest serves as a grouping variable for each multigroup analysis, and intervention status is included as a predictor of the SEC growth trajectory, along with all the other student-level covariates. The SEC growth trajectory comprises two latent factors, intercept (baseline level) and slope (growth rate), assuming linear growth throughout the year. A more flexible approach to estimating the shape of growth was limited due to the lack of degrees of freedom. The time coefficients of the slope factor were fixed at 0, 0.5, and 1, respectively, so that the estimated mean slope is interpreted as an average rate of growth from baseline to year-end. The Wald chi-square test of parameter equality was conducted in each multigroup analysis in order to examine whether the intervention effect (indicated by the regression coefficient of the slope factor on intervention status) significantly differs across student sociocultural subgroups of interest.

Next, in the second approach, a multigroup analysis by intervention status was conducted in order to understand the effects of SEL intervention on subgroup disparities in

SEC growth trajectories more straightforwardly by parameterizing subgroup differences in baseline and growth rate. In this approach, intervention status now serves as a grouping variable for multigroup analysis, and all the student-level characteristics are included as a set of predictors of the SEC growth trajectory (assumed to be linear as specified above). This model is presented in Figure 3.1. A set of Wald chi-square tests of parameter equality between the two conditions were conducted in order to examine whether the values of all the regression coefficients significantly differ between the two conditions.

These two modeling approaches, in essence, answer the same research question. However, the first approach is focused on parameterizing intervention effects and examining whether the effects were heterogeneous across student subgroups, which is a more comparable way of conducting "subgroup analysis" as in conventional intervention research (Bloom & Michalopoulos, 2013; Wang & Ware, 2013). On the other hand, the second approach is focused on parameterizing subgroup disparities in SEC growth trajectories and examining whether a universal SEL intervention had any effect on the extent to which these disparities change over time. To my knowledge, no study has been published using this approach to examine the effects of an intervention on subgroup disparities in the literature on SEL or more broadly preventive interventions for youth.

Figure 3.1. Multigroup Latent Growth Model of Student Social-Emotional Competence (SEC)



The model fit of the growth model was assessed based on the chi-square statistic as well as the comparative fit index (CFI), the Tucker-Lewis index (TLI), the standardized root mean squared residual (SRMR), and the root mean squared error of approximation (RMSEA). Given the sample size and model complexity, a CFI higher than .92, a TLI higher than .92, an

SRMR lower than .08, and an RMSEA lower than .07 indicate an acceptable fit, while significant p-values are expected for the chi-square tests (Hair et al., 2010). For the Wald tests of parameter equality, an alpha level of .05 was used as a threshold for statistical significance, with p-values between .05 and .10 reported as meeting the threshold for "marginal significance," given that such interaction effects are usually difficult to demonstrate due to limited statistical power (Bloom & Michalopoulos, 2013). All the main analyses were conducted using Mplus version 8 (Muthén & Muthén, 1998-2017). The Full Information Maximum Likelihood (FIML) estimator with robust standard errors (MLR in Mplus) was used to handle missingness and any violations to normality and independence assumptions.

Results

Descriptive Statistics

In the comparison condition, average student SEC scores increased by 1.37 T-Score points from Fall to Spring, whereas in the intervention condition, they increased by 3.41 T-Score points during the study year. The distribution of student SEC ratings was slightly right-skewed and platykurtic across all measurement time points. The absolute values of skewness and kurtosis were smaller than 1 for all items, however, which meets the customary thresholds for skewness and kurtosis (3 and 8, respectively; Kline, 2015; See Table 3.2). Around 70% of cases had complete data for all three outcome variables. The probability of having any missing data did not significantly differ by student gender, race and ethnicity, SES, disability, or ELL. Missingness was significantly greater in the comparison group ($\chi^2(1) = 11.54$, $p < .001$) but the difference was negligible (Phi = 0.085).

Table 3.2. Descriptive Statistics of DESSA-Mini T-Scores

Time	Valid N	M	SD	Skewness	Kurtosis	Cronbach's α
Comparison Condition						
Fall	521	50.25	10.79	0.28	-0.62	0.94
Winter	381	49.81	11.33	0.22	-0.73	0.95
Spring	461	51.62	11.29	0.31	-0.74	0.96
Intervention Condition						
Fall	941	50.58	11.13	0.22	-0.66	0.95
Winter	964	51.49	11.19	0.15	-0.67	0.96
Spring	912	53.99	11.76	0.05	-0.97	0.97

Comparing Intervention Effects Across Student Subgroups

All the multigroup analyses, separately done by gender, race and ethnicity, SES, disability status, and ELL status, showed acceptable model fit. The Wald test results from these analyses suggest that the intervention effects were not different across any of the examined subgroups, except for Black students. Specifically, the TOOLBOX intervention effects on the rate of SEC growth were significantly larger for Black students than White students ($\chi^2(1) = 4.08$, $p < .05$). When the intervention effects were compared between Black and all the other non-Black subgroups combined (including Asian, Hispanic, and White students), the intervention effects were still marginally larger for Black students ($\chi^2(1) = 3.56$, $p = .059$). No significant differences in the intervention effects were observed among the other three racial-ethnic subgroups.

Disparities in SEC Growth Trajectories Across Intervention Conditions

The multigroup analysis by intervention status also showed an acceptable fit ($\chi^2(18) = 77.83$, $p < .001$, CFI = 0.97, TLI = 0.92, RMSEA = 0.065, 90% CI [0.05, 0.08], SRMR = 0.02). Table 3.3 presents the regression coefficient estimates on the intercept and slope factors within each condition and the Wald test statistics associated with each parameter. As suggested by the previous modeling results, only the Black-White disparities in SEC growth trajectories showed a different pattern between the two intervention conditions among the examined subgroup disparities. At baseline, the gap between Black and White students was marginally larger for those under the intervention condition than the comparison condition ($\chi^2(1) = 3.82$, $p = .051$), suggesting that the baseline Black-White gap was wider in the intervention condition than in the comparison condition. Statistically controlling for these baseline differences, the SEC growth rates showed a marginally different pattern across intervention conditions ($\chi^2(1) = 3.08$, $p = .079$), suggesting that Black students had a higher rate of growth than White students in the intervention condition, more so than in the comparison condition.

Other than this Black-White gap, baseline disparities were observed by gender and disability status in both conditions (favoring female and non-SPED students). But the rates of SEC growth did not differ across gender or disability subgroups in either condition. No baseline subgroup difference was found across other racial-ethnic subgroups, SES, or ELL status, and any differences in their SEC growth rates did not differ by intervention condition at least at the $p < .10$ level.

Table 3.3. Estimation Results of the Conditional SEC Growth Model by Intervention Condition

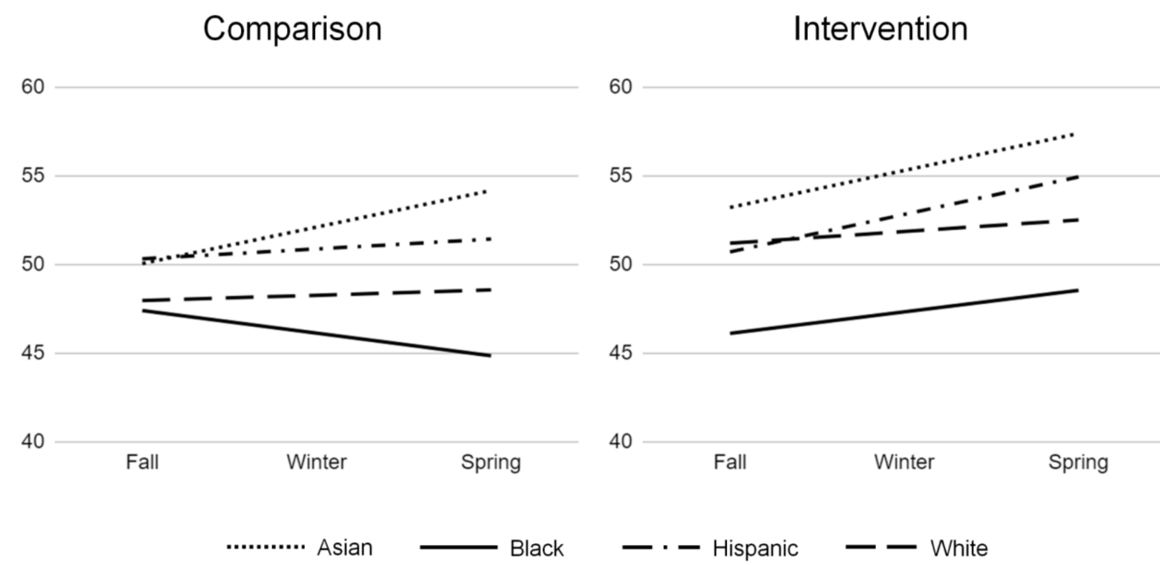
	Comparison (N=541)			Intervention (N=1049)			Wald Test	
	Est.	SE	*p*	Est.	SE	*p*	$\chi^2(1)$	*p*
Intercept								
Male	-3.22	0.88	<.001	-5.04	0.65	<.001	2.76	.100
Asian	2.09	1.75	.234	2.02	1.40	.150	0.00	.977
Black	-0.56	1.89	.770	-5.09	1.34	<.001	3.82	.051
Hispanic	2.36	1.47	.108	-0.49	1.17	.677	2.30	.129
Low-income	-0.28	0.98	.772	-0.46	0.74	.533	0.02	.885
SPED	-7.74	1.58	<.001	-8.91	1.09	<.001	0.38	.540
ELL	0.28	0.98	.779	-1.27	0.74	.085	1.58	.208
Grade	2.00	0.53	<.001	0.10	0.39	.806	8.46	.004

Table 3.3 (continued).

Slope								
Male	-0.81	0.83	.328	-0.76	0.66	.243	0.00	.962
Asian	3.52	1.90	.064	2.86	1.31	.029	0.08	.777
Black	-3.16	2.07	.127	1.11	1.28	.385	3.08	.079
Hispanic	0.51	1.64	.756	1.81	1.09	.095	0.44	.509
Low-income	-0.03	1.10	.979	-0.96	0.73	.184	0.51	.477
SPED	-2.14	1.51	.156	0.70	0.99	.477	2.48	.115
ELL	0.43	0.97	.655	1.79	0.75	.017	1.23	.267
Grade	-0.73	0.48	.128	-0.50	0.40	.205	0.14	.712

To facilitate interpretation of findings related to Black-White disparities, Figure 3.2 shows the estimated marginal means of student SEC over time across four racial-ethnic subgroups within each condition. Descriptively speaking, within the comparison condition, Black students' SEC showed a decreasing pattern over time, unlike White or other racial-ethnic subgroups. Within the intervention condition, however, Black students' SEC showed an increasing pattern over time, similar to White or other racial-ethnic subgroups.

Figure 3.2. Marginal Means of SEC Over Time Across Racial-Ethnic Subgroups Within Each Condition



Discussion

This study aimed to examine whether a universal social and emotional learning (SEL) intervention works similarly or differentially across diverse subgroups of students in order to understand its implications for subgroup disparities in social-emotional competence (SEC). Using data collected from a quasi-experimental study of the TOOLBOX program, as an example, this study examined the extent to which one-year implementation of universal SEL was differentially related to K-2 students' SEC growth across diverse student subgroups classified by gender, race and ethnicity, socioeconomic status, disability status, and English learner status. The results suggest that the intervention effects were slightly larger for Black students, compared to those for White or other racial-ethnic subgroups. In the comparison condition, the SEC disparities between Black and White students tended to widen throughout the year, whereas in the intervention condition, Black students showed a similar rate of growth as their White peers. In the meantime, the intervention effects were not different across any other student subgroups. In other words, the intervention did not have significant moderation effects on any other subgroup disparities in SEC growth rates. Implications of these findings and recommendations for future SEL research and practice are discussed as follows.

At the intersection of prevention and education fields, universal SEL has often been advocated as a promising approach to ensuring healthy development for *all* youth (Hawkins et al., 2015) and providing *equitable* access to a well-rounded education (U.S. Department of Education, Office of Elementary and Secondary Education, 2016). The efforts to facilitate widespread dissemination of SEL have been based on the assumption that it could reach and benefit all students, including disadvantaged students who would have had less opportunities to develop SEC without such an intervention (Assessment Work Group, 2019; Domitrovich et al., 2017; Elias & Haynes, 2008). However, as reviewed earlier, empirical evidence on whether universal SEL works equally well across diverse sociocultural subgroups of students has been limited and inconsistent. The findings of this study provide a piece of evidence illustrating that a universal SEL program might have comparable effects across many student subgroups and potentially have larger effects among some marginalized subgroups, in our case Black students.

In addition to disaggregating intervention effects across subgroups, this study also makes a unique contribution to the literature by examining the effects of universal SEL on subgroup disparities in SEC growth trajectory: universal SEL might prevent some disparities from further widening, but not necessarily reduce the existing subgroup disparities in SEC observed at baseline. Taken together, and if widely replicated, this evidence favors a public health approach to SEL that provides equal opportunities for all students to develop essential protective factors for their mental, emotional, and behavioral health; but at the same time, suggests an SEL curriculum may not be *sufficient* to achieve the goal of reducing subgroup disparities in social-emotional outcomes.

This study also raises interesting questions about the meaning of these findings and the implications for more equitable SEL practices and student outcomes. Although the field of SEL has been characterized as "intellectually diverse and politically nonpartisan" (Osher et al., 2016, p. 647), some scholars have criticized that the predominant practices of school-based SEL reflect hegemonic white, middle-class values, thus culturally misaligned for students from marginalized backgrounds (Gregory & Fergus, 2017; Hoffman, 2009; Pyscher & Crampton, 2020). From this view, the present findings of larger effects among Black students could be suggestive of assimilation rather than positive development: Black students might have learned to conform to the social-emotional norms and expectations of the dominant school culture, likely at the expense of their own cultural identity and affirmation. If this was the case, the goal of reducing disparities may conflict with the goal of promoting equity.

In light of this possibility, there has recently been a discussion of *Transformative SEL* (T-SEL; Collaborative for Academic, Social, and Emotional Learning, 2022). T-SEL promotes focal constructs such as the development of identity, agency, and belonging, as well as the co-creation of an inclusive and engaging learning environment to promote equity and justice (Jagers et al., 2018, 2019). From this view, SEL can be constructed and implemented in a way that celebrates diversity and disrupts existing inequities in the school system, achieving the dual goals of promoting individual well-being and equity (e.g., Lea et al., 2021; Ozer et al., 2020).

As for limitations, the data analyzed in this study do not provide any further information to explain, for example, *why* the intervention had larger effects for Black students. It should also be noted that this study cannot claim strong causal effects due to lack of randomization, and it has limited generalizability beyond the study sample and contexts. Still, through an example of a universal SEL program implemented in routine school settings, this study calls attention to the important but understudied questions about disparities and equity in the field of SEL. By explicitly testing the subgroup differences in the intervention effects, this study emphasizes the need to answer the question of *for whom* universal SEL works, by not assuming homogeneous effects for all and only reporting average effects. Also, by comparing the patterns of subgroup disparities in SEC growth with and without an SEL intervention, this study raises an important question as to the role of universal SEL in disparity reduction and equity enhancement. Admittedly, much more work is needed to fully explore the mechanisms and contexts that explain observed disparities and differential intervention effects across sociocultural subgroups (e.g., Black students having qualitatively different experiences at schools; Kim et al., 2022) and the best strategies for school-based preventive interventions to produce strong and equitable social-emotional outcomes.

Conclusion

This dissertation aimed to provide empirical evidence that informs important but under-investigated questions in the literature on social and emotional learning (SEL). In Chapter 1, I examined how student social-emotional competence (SEC) develops, on average, across elementary school years under a routine SEL practice condition, after first testing the longitudinal measurement invariance of a widely-used SEC monitoring tool. The growth modeling results suggest that student SEC increased within each year and decreased over each summer by about a half of the yearly gain, while the rate of yearly growth gradually decreased across years. In Chapter 2, I examined the extent to which disparities exist in student SEC developmental trajectories across socioculturally classified subgroups, after first testing the measurement invariance of a SEC progress monitoring tool across student gender, race and ethnicity, and socioeconomic status. The growth modeling results suggest that male (vs. female), Black or Hispanic (vs. White), and lower-income (vs. higher-income) students started with a lower level of SEC, and these gaps were either sustained across time or slightly widened. In Chapter 3, I examined the similarities and differences in the intervention effects of universal SEL across diverse student subgroups, and the effects of universal SEL on existing subgroup disparities. The multigroup growth modeling results suggest that the intervention effects were similar across student subgroups classified by gender, socioeconomic status, disability, and English language learning status, while the effects were slightly larger for Black students, compared to White or other racial-ethnic subgroups. Implications and limitations of each of these findings were discussed with suggestions for future research and practices.

In summary, this dissertation contributes to the literature by providing evidence of (1) an average multi-year developmental trajectory of SEC among school-aged children, assessed with a measure tested to be longitudinally invariant, (2) gender, racial-ethnic, and socioeconomic disparities that exist in this SEC developmental trajectory, assessed with a measure tested to be invariant across subgroups, and (3) the potential of universal SEL for preventing some disparities from further widening, but not necessarily closing all the existing gaps. I hope the findings of this dissertation lead to and inform larger questions in the field, such as how to make sure SEL assessment tools are accurate and equitable, how to set SEL standards and benchmarks, and how to ensure school-based SEL truly benefits all students and produces more equitable outcomes.

References

Aber, J. L., Brown, J. L., & Jones, S. M. (2003). Developmental trajectories toward violence in middle childhood: Course, demographic differences, and response to school-based intervention. *Developmental Psychology*, *39*(2), 324–348. http://dx.doi.org/10.1037/0012-1649.39.2.324

Alegría, M., Green, J. G., McLaughlin, K. A., & Loder, S. (2015). *Disparities in child and adolescent mental health and mental health services in the US*. William T. Grant Foundation.

Alvidrez, J., Castille, D., Laude-Sharp, M., Rosario, A., & Tabor, D. (2019). The national institute on minority health and health disparities research framework. *American Journal of Public Health*, *109*(S1), S16–S20.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association.

Anderson-Butcher, D., Amorose, A. J., Lower, L. M., Riley, A., Gibson, A., & Ruch, D. (2016). The case for the perceived social competence scale II. *Research on Social Work Practice*, *26*(4), 419–428.

Anthony, C. J., Elliott, S. N., DiPerna, J. C., & Lei, P.-W. (2020). Initial development and validation of the Social Skills Improvement System—Social and Emotional Learning Brief Scales-Teacher Form. *Journal of Psychoeducational Assessment*, *0*(0), 1–16. https://doi.org/10.1177/0734282920953240

Ardoin, S. P., Witt, J. C., Connell, J. E., & Koenig, J. L. (2005). Application of a three-tiered response to intervention model for instructional planning, decision making, and the identification of children in need of services. *Journal of Psychoeducational Assessment*, *23*(4), 362–380. https://doi.org/10.1177/073428290502300405

Arunkumar, R., Midgley, C., & Urdan, T. (1999). Perceiving high or low home-school dissonance: Longitudinal effects on adolescent emotional and academic well-being. *Journal of Research on Adolescence*, *9*(4), 441–466.

Assessment Work Group. (2019). *Student Social and Emotional Competence Assessment: The Current State of the Field and a Vision for Its Future*. CASEL.

Assessment Work Group. (2021). *SEL Assessment Guide*. Measuring SEL. https://measuringsel.casel.org/assessment-guide/

Bauer, G. R. (2014). Incorporating intersectionality theory into population health research methodology: Challenges and the potential to advance health equity. *Social Science & Medicine*, *110*, 10–17.

Bierman, K. L., Coie, J. D., Dodge, K. A., Greenberg, M. T., Lochman, J. E., McMahon, R. J., & Pinderhughes, E. (2010). The effects of a multiyear universal social–emotional learning program: The role of student and school characteristics. *Journal of Consulting and Clinical Psychology*, *78*(2), 156. https://doi.org/10.1037/a0018607

Bloom, H. S., & Michalopoulos, C. (2013). When is the story in the subgroups? *Prevention Science*, *14*(2), 179–188.

Borst, G., & Srinivasan, N. (2022). Understanding the Social and Emotional Brain. In *S. N.*

Chatterjee & A. K. Duraiappah. (Eds.). Rethingking Learning: A Review of Social and Emotional Learning for Education Systems (pp. 31–52). United Nations Educational, Scientific and Cultural Organization, Mahatma Gandhi Institute of Education for Peace and Sustainable Development (UNESCO MGIEP).

Bowleg, L. (2012). The problem with the phrase women and minorities: Intersectionality—An important theoretical framework for public health. *American Journal of Public Health*, *102*(7), 1267–1273.

Boykin, A. W., Tyler, K. M., & Miller, O. (2005). In search of cultural themes and their expressions in the dynamics of classroom life. *Urban Education*, *40*(5), 521–549.

Bradshaw, C. P., Mitchell, M. M., O'Brennan, L. M., & Leaf, P. J. (2010). Multilevel exploration of factors contributing to the overrepresentation of black students in office disciplinary referrals. *Journal of Educational Psychology*, *102*(2), 508–520.

Buckley, M., & Saarni, C. (2009). Emotion regulation: Implications for positive youth development. In *R. Gilman, E. S. Huebner, & M. J. Furlong (Eds.) Handbook of Positive Psychology in Schools*. Routledge.

Carroll, A., Houghton, S., Forrest, K., McCarthy, M., & Sanders-O'Connor, E. (2020). Who benefits most? Predicting the effectiveness of a social and emotional learning intervention according to children's emotional and behavioural difficulties. *School Psychology International*, *41*(3), 197–217.

Catalano, R. F., Berglund, M. L., Ryan, J. A., Lonczak, H. S., & Hawkins, J. D. (2002). Positive youth development in the United States: Research findings on evaluations of positive youth development programs. *Prevention & Treatment*, *5*(15).

Chain, J., Shapiro, V. B., LeBuffe, P. A., & Bryson, A. M. (2017). Academic achievement of American Indian and Alaska native students: Does social-emotional competence reduce the impact of poverty? *American Indian and Alaska Native Mental Health Research*, *24*(1), 1–29.

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, *95*(5), 1005–1018.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255.

Cohen, J. (1969). *Statistical Power Analysis for the Behavioral Sciences* (1st ed). Academic Press.

Collaborative for Academic, Social, and Emotional Learning. (2020). *CASEL's SEL Framework: What Are the Core Competence Areas and Where Are They Promoted?* https://casel.org/wp-content/uploads/2020/12/CASEL-SEL-Framework-11.2020.pdf

Collaborative for Academic, Social, and Emotional Learning. (2022a). *Transformative SEL*. Collaborative for Academic, Social, and Emotional Learning. https://casel.org/fundamentals-of-sel/how-does-sel-support-educational-equity-and-excellence/transformative-sel/

Collaborative for Academic, Social, and Emotional Learning. (2022b). *What Is the CASEL Framework?* Collaborative for Academic, Social, and Emotional Learning. https://casel.org/fundamentals-of-sel/what-is-the-casel-framework/

Colle, L., & Del Giudice, M. (2011). Patterns of attachment and emotional competence in middle childhood. *Social Development*, *20*(1), 51–72.
Collins, M. A. (2015). *TOOLBOX^{TM} Primer*. Dovetail Learning, Inc.
Collins, W. A. (1984). *Development during middle childhood: The years from six to twelve*. National Academy Press.
Conduct Problems Prevention Research Group. (1999). Initial impact of the Fast Track prevention trial for conduct problems: II. Classroom effects. *Journal of Consulting and Clinical Psychology*, *67*(5), 648.
Corcoran, R. P., Cheung, A. C., Kim, E., & Xie, C. (2018). Effective universal school-based social and emotional learning programs for improving academic achievement: A systematic review and meta-analysis of 50 years of research. *Educational Research Review*, *25*, 56–72.
Côté, S., Tremblay, R. E., Nagin, D., Zoccolillo, M., & Vitaro, F. (2002). The development of impulsivity, fearfulness, and helpfulness during childhood: Patterns of consistency and change in the trajectories of boys and girls. *Journal of Child Psychology and Psychiatry*, *43*(5), 609–618.
Denham, S. A. (2006). Social-emotional competence as support for school readiness: What is it and how do we assess it? *Early Education and Development*, *17*(1), 57–89. https://doi.org/10.1207/s15566935eed1701_4
Denham, S. A., Bassett, H. H., & Wyatt, T. M. (2010). Gender differences in the socialization of preschoolers' emotional competence. *New Directions for Child and Adolescent Development*, *2010*(128), 29–49.
District. (2016). *Data Dashboard*. [Masked URL]
Dixon-Román, E. J. (2017). *Inheriting Possibility: Social Reproduction and Quantification in Education*. The University of Minnesota Press.
Domitrovich, C. E., Durlak, J. A., Staley, K. C., & Weissberg, R. P. (2017). Social-emotional competence: An essential factor for promoting positive adjustment and reducing risk in school children. *Child Development*, *88*(2), 408–416. https://doi.org/10.1111/cdev.12739
Duchesneau, N. (2020). *Social, Emotional, and Academic Development through an Equity Lens.* The Education Trust. https://files.eric.ed.gov/fulltext/ED607298.pdf
Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, *44*(4), 237–251. https://doi.org/10.3102/0013189X15584327
Duncan, R., Washburn, I. J., Lewis, K. M., Bavarian, N., DuBois, D. L., Acock, A. C., Vuchinich, S., & Flay, B. R. (2017). Can universal SEL programs benefit universally? Effects of the Positive Action program on multiple trajectories of social-emotional and misconduct behaviors. *Prevention Science*, *18*(2), 214–224. https://doi.org/10.1007/s11121-016-0745-1
Duncan, T. E., & Duncan, S. C. (2004). An introduction to latent growth curve modeling. *Behavior Therapy*, *35*(2), 333–363.
Durlak, J. A., & Weissberg, R. P. (2011). Promoting social and emotional development is an essential part of students' education. *Human Development*, *54*(1), 1–3.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, *82*(1), 405–432. https://doi.org/10.1111/j.1467-8624.2010.01564.x

Eisenberg, N., Fabes, R. A., & Spinrad, T. L. (2006). Prosocial development. In *N. Eisenberg (Ed.), Handbook of Child Psychology* (Vol. 1–3, pp. 646–718). John Wiley & Sons. http://onlinelibrary.wiley.com/doi/10.1002/9780470147658.chpsy0311/full

El Mallah, S. (2020). Toward equity-oriented assessment of social and emotional learning: Examining equivalence of concepts and measures. *Urban Education*.

Elias, M. J., & Haynes, N. M. (2008). Social competence, social support, and academic achievement in minority, low-income, urban elementary school children. *School Psychology Quarterly*, *23*(4), 474–495. https://doi.org/10.1037/1045-3830.23.4.474

Elias, M. J., Leverett, L., Duffell, J. C., Humphrey, N., Stepney, C., & Ferrito, J. (2015). Integrating SEL with related prevention and youth development approaches. In *J. A. Durlak, C. E. Domitrovich, R. P. Weissberg, & T. P. Gullotta (Eds.), Handbook of Social and Emotional learning: Research and Practice* (pp. 33–49). Guilford Publications.

Elias, M. J., Zins, J. E., Weissberg, R. P., Frey, K. S., Greenberg, M. T., Haynes, N. M., Kessler, R., Schwab-Stone, M. E., & Shriver, T. P. (1997). *Promoting social and emotional learning: Guidelines for educators*. Association for Supervision and Curriculum Development.

Farrell, A. D., & Vulin-Reynolds, V. (2007). Violent behavior and the science of prevention. In *In Flannery, D. J., Vazsonyi, A. T., & Waldman, I. D. (Eds.). The Cambridge Handbook of Violent Behavior and Aggression* (pp. 19–48). Cambridge University Press.

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*(5), 532.

Ferrer, E., Balluerka, N., & Widaman, K. F. (2008). Factorial invariance and the specification of second-order latent growth models. *Methodology*, *4*(1), 22–36.

Fishbein, D. H., Domitrovich, C., Williams, J., Gitukui, S., Guthrie, C., Shapiro, D., & Greenberg, M. (2016). Short-term intervention effects of the PATHS curriculum in young low-income children: Capitalizing on plasticity. *The Journal of Primary Prevention*, *37*(6), 493–511.

Fleming, J. L. (2014). *Allentown School District Social-Emotional Learning (SEL) Initiative: Final Year 1 Evaluation Report 2011-2012*. The Devereux Center for Resilient Children.

Fredrick, S. S., Drevon, D. D., & Jervinsky, M. (2019). Measurement invariance of the Student Risk Screening Scale across time and gender. *School Psychology*, *34*(2), 159–167.

Frey, K. S., Hirschstein, M. K., Snell, J. L., Edstrom, L. V. S., MacKenzie, E. P., & Broderick, C. J. (2005). Reducing playground bullying and supporting beliefs: An experimental trial of the steps to respect program. *Developmental Psychology*, *41*(3), 479–490.

Frey, K. S., Nolen, S. B., Edstrom, L. V. S., & Hirschstein, M. K. (2005). Effects of a school-based social–emotional competence program: Linking children's goals, attributions, and behavior. *Journal of Applied Developmental Psychology*, *26*(2), 171–200.

Frohlich, K. L., & Potvin, L. (2008). The inequality paradox: The population approach and vulnerable populations. *American Journal of Public Health*, *98*(2), 216–221.

Garner, P. W., Mahatmya, D., Brown, E. L., & Vesely, C. K. (2014). Promoting desirable outcomes among culturally and ethnically diverse children in social emotional learning programs: A multilevel heuristic model. *Educational Psychology Review*, *26*(1), 165–189.

Gayl, C. L. (2017). *How State Planning for the Every Student Succeeds Act (ESSA) Can Promote Student Academic, Social, and Emotional Learning: An Examination of Five Key Strategies*. Collaborative for Academic, Social, and Emotional Learning (CASEL).

Gehlbach, H., & Hough, H. J. (2018). *Measuring Social Emotional Learning through Student Surveys in the CORE Districts: A Pragmatic Approach to Validity and Reliability.* Policy Analysis for California Education, PACE. https://files.eric.ed.gov/fulltext/ED591082.pdf

Gottfredson, D. C., Cook, T. D., Gardner, F. E., Gorman-Smith, D., Howe, G. W., Sandler, I. N., & Zafft, K. M. (2015). Standards of evidence for efficacy, effectiveness, and scale-up research in prevention science: Next generation. *Prevention Science*, *16*(7), 893–926. https://doi.org/10.1007/s11121-015-0555-x

Grant, S., Hamilton, L. S., Wrabel, S. L., Gomez, C. J., Whitaker, A., Leschitz, J. T., Unlu, F., Chavez-Herrerias, E. R., Baker, G., & Barrett, M. (2017). *How the Every Student Succeeds Act Can Support Social and Emotional Learning.* RAND Corporation.

Greenberg, M. T., Domitrovich, C. E., Weissberg, R. P., & Durlak, J. A. (2017). Social and emotional learning as a public health approach to education. *The Future of Children*, 13–32.

Greenberg, M. T., Kusche, C. A., Cook, E. T., & Quamma, J. P. (1995). Promoting emotional competence in school-aged children: The effects of the. *Development and Psychopathology*, *7*, 117–136.

Gregory, A., & Fergus, E. (2017). Social and emotional learning and equity in school discipline. *The Future of Children*, *27*(1), 117–136.

Gregory, A., & Roberts, G. (2017). Teacher beliefs and the overrepresentation of Black students in classroom discipline. *Theory Into Practice*, *56*(3), 187–194.

Gresham, F. M., Cook, C. R., Collins, T., Dart, E., Rasetshwane, K., Truelson, E., & Grant, S. (2010). Developing a change-sensitive brief behavior rating scale as a progress monitoring tool for social behavior: An example using the Social Skills Rating System—Teacher Form. *School Psychology Review*, *39*(3), 364–379.

Hair, J. K., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Pearson.

Hamilton, L. S., & Doss, C. J. (2020). *Supports for Social and Emotional Learning in American Schools and Classrooms*. RAND Corporation.

Hawkins, J. D., Jenson, J. M., Catalano, R. F., Fraser, M. W., Botvin, G. J., Shapiro, V. B., the, & the Coalition for Behavioral Health. (2015). *Unleashing the power of prevention (Discussion Paper)*. Washington, DC: Institute of Medicine and National Research Council.

Hay, D. F., Paine, A. L., Perra, O., Cook, K. V., Hashmi, S., Robinson, C., Kairis, V., & Slade, R. (2021). Prosocial and aggressive behavior: A longitudinal study. *Monographs of the Society for Research in Child Development*, *86*(2), 7–103.

Hoffman, D. M. (2009). Reflecting on social emotional learning: A critical perspective on trends in the United States. *Review of Educational Research*, *79*(2), 533–556.

Holsen, I., Iversen, A. C., & Smith, B. H. (2009). Universal social competence promotion programme in school: Does it work for children with low socio-economic background? *Advances in School Mental Health Promotion*, *2*(2), 51–60. https://doi.org/10.1080/1754730X.2009.9715704

Hutchison, M., Russell, B. S., & Wink, M. N. (2020). Social-emotional competence trajectories from a school-based child trauma symptom intervention in a disadvantaged community. *Psychology in the Schools*, *57*(8), 1257–1272.

Institute of Education Sciences. (2020). *Characteristics of Public and Private Elementary and Secondary School Teachers in the United States: Results From the 2017–18 National Teacher and Principal Survey* (NCES 2020-142rev). The National Center for Education Statistics, U.S. Department of Education. https://nces.ed.gov/pubs2020/2020142rev.pdf

Jagers, R. J., Rivas-Drake, D., & Borowski, T. (2018). *Equity and social and emotional learning: A cultural analysis* (The Measuring SEL Series of Frameworks Briefs). The Assessment Work Group. https://measuringsel.casel.org/wp-content/uploads/2018/11/Frameworks-Equity.pdf

Jagers, R. J., Rivas-Drake, D., & Williams, B. (2019). Transformative social and emotional learning (SEL): Toward SEL in service of educational equity and excellence. *Educational Psychologist*, *54*(3), 162–184.

Jones, D. J., Greenberg, M. T., & Crowley, M. (2015). The economic case for SEL. In *J. A. Durlak, C. E. Domitrovich, R. P. Weissberg, & T. P. Gullotta (Eds.), Handbook of Social and Emotional learning: Research and Practice* (pp. 97–113). Guilford Publications.

Jones, S. M., Brown, J. L., & Aber, J. L. (2011). Two-year impacts of a universal school-based social-emotional and literacy intervention: An experiment in translational developmental research. *Child Development*, *82*(2), 533–554.

Jones, T. M. (2018). *Understanding race differences in academic outcomes, school climate and social emotional learning to promote racial equity and policy reform* [Doctoral dissertation]. University of Washington.

Jones, T. M., Fleming, C., & Williford, A. (2020). Racial equity in academic success: The role of school climate and social emotional learning. *Children and Youth Services Review*, *119*. https://doi.org/10.1016/j.childyouth.2020.105623

Kautz, T., Heckman, J. J., Diris, R., Ter Weel, B., & Borghans, L. (2014). *Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success* (No. w20749). National Bureau of Economic Research.

Kilbourne, A. M., Switzer, G., Hyman, K., Crowley-Matoka, M., & Fine, M. J. (2006). Advancing health disparities research within the health care system: A conceptual framework. *American Journal of Public Health*, *96*(12), 2113–2121.

Kim, B. K. E., De Leon, J., Quinn, C. R., Logan-Greene, P. B., & Nurius, P. S. (2022). Understanding the School-to-Prison Pipeline for Black Probation Youth. In *P. Burch (Ed.), System Failure: Policy and Practice in the School to Prison Pipeline* (pp. 100–117). Routledge/Taylor and Francis.

Kim, B. K. E., Oesterle, S., Hawkins, J. D., & Shapiro, V. B. (2015). Assessing sustained effects of Communities That Care on youth protective factors. *Journal of Society for Social Work and Research*, *6*(4), 565–589.

Kim, E. S., & Willson, V. L. (2014). Measurement invariance across groups in latent growth modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(3), 408–424.

King, E. K. (2021). Fostering toddlers' social emotional competence: Considerations of teachers' emotion language by child gender. *Early Child Development and Care*, *191*(16), 2494–2507.

Kline, R. B. (2015). *Principles and Practice of Structural Equation Modeling* (Fourth Edition). Guilford publications.

Koch, T., Holtmann, J., Eid, M., & West, S. G. (2020). Analyzing longitudinal multirater data with changing and stable raters. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(1), 73–87.

Kokko, K., Tremblay, R. E., Lacourse, E., Nagin, D. S., & Vitaro, F. (2006). Trajectories of prosocial behavior and physical aggression in middle childhood: Links to adolescent school dropout and physical violence. *Journal of Research on Adolescence*, *16*(3), 403–428.

Krachman, S. B., & LaRocca, B. (2017). *The Scale of Our Investment in Social-Emotional Learning.* Transforming Education. https://transformingeducation.org/wp-content/uploads/2017/10/Inspire-Paper-Transforming-Ed-FINAL-2.pdf

Kraft, M. A. (2019). *Interpreting Effect Sizes of Education Interventions* (EdWorkingPaper No. 19-10). Annenberg Institute at Brown University. https://scholar.harvard.edu/files/mkraft/files/kraft_2019_effect_sizes.pdf

Krishnan, V. (2011). Teachers' assessment of preschoolers' social and emotional competence: Does sex of children matter in developmental outcomes? *Proceedings of the World Academy of Science, Engineering and Technology, Year 7, Issue 79*, 1622–1632.

Kusché, C. A., & Greenberg, M. T. (1994). *The PATHS Curriculum: Promoting Alternative Thinking Strategies*. Developmental Research & Programs.

Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in US schools. *Educational Researcher*, *35*(7), 3–12.

Lareau, A. (2011). *Unequal Childhoods: Class, Race, and Family Life, with an Update a Decade Later*. The University of California Press.

Lea, C. H., McCowan, K. J., Jones, T. M., & Malorni, A. (2021). Adult and student perspectives on racial and ethnic equity-informed school-based strategies. *Psychology in the Schools*, 1–20. https://doi.org/10.1002/pits.22575

LeBuffe, P. A., Shapiro, V. B., & Naglieri, J. A. (2009). *The Devereux Student Strengths Assessment (DESSA) Assessment, Technical Manual, and User's Guide*. Apperson Inc.

LeBuffe, P. A., Shapiro, V. B., & Robitaille, J. L. (2018). The Devereux Student Strengths Assessment (DESSA) comprehensive system: Screening, assessing, planning, and monitoring. *Journal of Applied Developmental Psychology*, *55*, 62–70.

Lee, J., Shapiro, V. B., Kim, B. E., & Yoo, J. P. (2018). Multilevel structural equation modeling for social work researchers: An introduction and application to healthy youth

development. *Journal of the Society for Social Work and Research*, *9*(4), 689–719.
Lee, J., Shapiro, V. B., Robitaille, J. L., & LeBuffe, P. A. (2022a). Gender, racial-ethnic, and socioeconomic disparities in the development of social-emotional competence among elementary school students. *Manuscript Submitted for Publication*.
Lee, J., Shapiro, V. B., Robitaille, J. L., & LeBuffe, P. A. (2022b). Measuring the development of social-emotional competence in elementary school years using behavioral rating scales. *Manuscript Submitted for Publication*.
Leonardo, Z. (2004). The color of supremacy: Beyond the discourse of 'white privilege.' *Educational Philosophy and Theory*, *36*(2), 137–152.
Leonardo, Z. (2012). The race for class: Reflections on a critical raceclass theory of education. *Educational Studies*, *48*(5), 427–449. https://doi.org/10.1080/00131946.2012.715831
Lewallen, T. C., Hunt, H., Potts-Datema, W., Zaza, S., & Giles, W. (2015). The Whole School, Whole Community, Whole Child model: A new approach for improving educational attainment and healthy development for students. *Journal of School Health*, *85*(11), 729–739.
Lorenc, T., Petticrew, M., Welch, V., & Tugwell, P. (2013). What types of interventions generate inequalities? Evidence from systematic reviews. *Journal of Epidemiology & Community Health*, *67*(2), 190–193.
Low, S., Smolkowski, K., Cook, C., & Desfosses, D. (2019). Two-year impact of a universal social-emotional learning curriculum: Group differences from developmentally sensitive trends over time. *Developmental Psychology*, *55*(2), 415–433.
Mahfouz, J., & Anthony-Stevens, V. (2020). Why trouble SEL? The need for cultural relevance in SEL. *Occasional Paper Series*, *2020*(43), 58–70.
Mahoney, J. L., LeBuffe, P. A., Shapiro, V. B., Robitaille, J. L., Johnson, E. S., & Adamson, J. L. (2022). *What is assessment bias and how is Aperture Education working to reduce it? (Ed. 1)*. Aperture Education.
Malti, T., Ribeaud, D., & Eisner, M. P. (2011). The effectiveness of two universal preventive interventions in reducing children's externalizing behavior: A cluster randomized controlled trial. *Journal of Clinical Child & Adolescent Psychology*, *40*(5), 677–692.
Mantz, L. S., Bear, G. G., Yang, C., & Harris, A. (2018). The delaware social-emotional competency scale (DSECS-S): Evidence of validity and reliability. *Child Indicators Research*, *11*(1), 137–157.
Mason, B. A., Gunersel, A. B., & Ney, E. A. (2014). Cultural and ethnic bias in teacher ratings of behavior: A criterion-focused review. *Psychology in the Schools*, *51*(10), 1017–1030.
McKown, C. (2019a). Challenges and opportunities in the applied assessment of student social and emotional learning. *Educational Psychologist*, *54*(3), 205–221. https://doi.org/10.1080/00461520.2019.1614446
McKown, C. (2019b). Reliability, factor structure, and measurement invariance of a web-based assessment of children's social-emotional comprehension. *Journal of Psychoeducational Assessment*, *37*(4), 435–449.
Melnick, H., Cook-Harvey, C. M., & Darling-Hammond, L. (2017). *Encouraging Social and Emotional Learning in the Context of New Accountability*. Learning Policy Institute.

Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care*, S69–S77.
Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*(1), 107–122.
Millsap, R. E. (2010). Testing measurement invariance using item response theory in longitudinal data: An introduction. *Child Development Perspectives*, *4*(1), 5–9. https://doi.org/10.1111/j.1750-8606.2009.00109.x
Millsap, R. E., & Everson, H. T. (1993). Methodology review: Statistical approaches for assessing measurement bias. *Applied Psychological Measurement*, *17*(4), 297–334.
Morsy, L., & Rothstein, R. (2019). Toxic Stress and Children's Outcomes: African American Children Growing up Poor Are at Greater Risk of Disrupted Physiological Functioning and Depressed Academic Achievement. *Economic Policy Institute*.
Muñoz, R. F., Mrazek, P. J., & Haggerty, R. J. (1996). Institute of Medicine report on prevention of mental disorders: Summary and commentary. *American Psychologist*, *51*(11), 1116–1122.
Muthén, L. K., & Muthén, B. O. (1998). *Mplus User's Guide.* (Eighth Edition). Muthén & Muthén.
Naglieri, J. A., LeBuffe, P. A., & Shapiro, V. B. (2011). *The Devereux Student Strengths Assessment Assessment—Mini (DESSA-Mini): Assessment, Technical Manual, and User's Guide*. Apperson Inc.
Naglieri, J. A., LeBuffe, P. A., & Shapiro, V. B. (2013). Assessment of social-emotional competencies related to resilience. In *S. Goldstein & R. B. Brooks (Eds.) Handbook of Resilience in Children* (2nd ed., pp. 261–272). Springer. https://doi.org/10.1007/978-1-4614-3661-4_15
National Academies of Sciences, Engineering, and Medicine. (2019). *Fostering healthy mental, emotional, and behavioral development in children and youth: A national agenda*. The National Academies Press.
National Research Council. (2012). *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*. The National Academies Press. https://doi.org/10.17226/13398
National Scientific Council on the Developing Child. (2007). *The science of early childhood development: Closing the gap between what we know and what we do*. Center on the Developing Child, Harvard University.
O'Brennan, L. M., Waasdorp, T. E., Pas, E. T., & Bradshaw, C. P. (2015). Peer victimization and social-emotional functioning: A longitudinal comparison of students in general and special education. *Remedial and Special Education*, *36*(5), 275–285.
Osher, D., Kidron, Y., Brackett, M., Dymnicki, A., Jones, S., & Weissberg, R. P. (2016). Advancing the science and practice of social and emotional learning: Looking back and moving forward. *Review of Research in Education*, *40*(1), 644–681. https://doi.org/10.3102/0091732X16673595
Ozer, E. J., Shapiro, V. B., & Duarte, C. d. P. (2020). *Opportunities to strengthen SEL impact through Youth-led Participatory Action Research (YPAR)*. Edna Bennett Pierce Prevention Research Center, The Pennsylvania State University.
Panorama Education. (2017). *Teacher Perception of Students' SEL: Using Panorama to collect*

teacher perception of students' social-emotional learning skills. panorama-www.s3.amazonaws.com
Peck, S. C., Smith, C., Hillaker, B., Macleod, C., Roy, L., Helegda, K., & Smith, L. (2018). *Measurement Manual for the Staff Rating of Youth Behavior-Short Form*.
Pendergast, L. L., von der Embse, N., Kilgus, S. P., & Eklund, K. R. (2017). Measurement equivalence: A non-technical primer on categorical multi-group confirmatory factor analysis in school psychology. *Journal of School Psychology*, *60*, 65–82.
Pennsylvania Department of Education. (2021). *Public School Enrollments*. https://www.education.pa.gov/DataAndReporting/Enrollment/Pages/PublicSchEnrReports.aspx
Perez, A. D., & Hirschman, C. (2009). The changing racial and ethnic composition of the US population: Emerging American identities. *Population and Development Review*, *35*(1), 1–51.
Perrino, T., Beardslee, W., Bernal, G., Brincks, A., Cruden, G., Howe, G., Murry, V., Pantin, H., Prado, G., & Sandler, I. (2015). Toward scientific equity for the prevention of depression and depressive symptoms in vulnerable youth. *Prevention Science*, *16*(5), 642–651.
Peters, C. D., Kranzler, J. H., Algina, J., Smith, S. W., & Daunic, A. P. (2014). Understanding disproportionate representation in special education by examining group differences in behavior ratings. *Psychology in the Schools*, *51*(5), 452–465.
Pitts, S. C., West, S. G., & Tein, J.-Y. (1996). Longitudinal measurement models in evaluation research: Examining stability and change. *Evaluation and Program Planning*, *19*(4), 333–350.
Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90.
Pyscher, T., & Crampton, A. (2020). Possibilities and problems in trauma-based and social emotional learning programs. *Occasional Paper Series*, *2020*(43), 3–9.
Raver, C. C., & Knitzer, J. (2002). *Ready to Enter: What Research Tells Policymakers About Strategies to Promote Social and Emotional School Readiness Among Three-and Four-Year-Old Children*. National Center for Children in Poverty, Columbia University.
Rodkin, P. C., Ryan, A. M., Jamison, R., & Wilson, T. (2013). Social goals, social behavior, and social status in middle childhood. *Developmental Psychology*, *49*(6), 1139–1150.
Rose-Krasnor, L., & Denham, S. (2009). Social-emotional competence in early childhood. In *K. H. Rubin, W. M. Bukowski, & B. P. Laursen (Eds.) Handbook of Peer Interactions, Relationships, and Groups* (pp. 162–179). Guilford Publications.
Ross, K. M., & Tolan, P. (2018). Social and emotional learning in adolescence: Testing the CASEL model in a normative sample. *The Journal of Early Adolescence*, *38*(8), 1170–1199.
Rowe, H. L., & Trickett, E. J. (2018). Student diversity representation and reporting in universal school-based social and emotional learning programs: Implications for generalizability. *Educational Psychology Review*, *30*(2), 559–583.
Ryan, R. M., Huta, V., & Deci, E. L. (2008). Living well: A self-determination theory

perspective on eudaimonia. *Journal of Happiness Studies*, *9*(1), 139–170. https://doi.org/10.1007/s10902-006-9023-4
Sarcedo, G. L., Matias, C. E., Montoya, R., & Nishi, N. (2015). Dirty dancing with race and class: Microaggressions toward first-generation and low income college students of color. *Journal of Critical Scholarship on Higher Education and Student Affairs*, *2*(1), 1.
Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*, *75*(2), 243–248.
Schonfeld, D. J., Adams, R. E., Fredstrom, B. K., Weissberg, R. P., Gilman, R., Voyce, C., Tomlin, R., & Speese-Linehan, D. (2015). Cluster-randomized trial demonstrating impact on academic achievement of elementary social-emotional learning. *School Psychology Quarterly*, *30*(3), 406.
Shapiro, V. B., Accomazzo, S., & Robitaille, J. L. (2017). In the same ballpark or a whole new ball game? Staff as raters of youth behavior. *Journal of Child and Family Studies*, *26*(4), 1051–1055. https://doi.org/10.1007/s10826-016-0632-1
Shapiro, V. B., Derr, A. S., Eldeeb, N., McCoy, H., Trujillo, M. A., & Vu, C. T. (2022). Unleashing the Power of Prevention to Ensure the Healthy Development for Youth: Expansions and Elaborations for Equity. In *M. Teasley & M. Spencer (Eds.), Racism and the Grand Challenges for the Social Work Profession*. Oxford University Press.
Shapiro, V. B., Kim, B. E., Robitaille, J. L., LeBuffe, P. A., & Ziemer, K. L. (2018). Efficient implementation monitoring in routine prevention practice: A grand challenge for schools. *Journal of the Society for Social Work and Research*, *9*(3), 377–394.
Shapiro, V. B., Kim, B. K. E., Accomazzo, S., & Roscoe, J. N. (2016). Predictors of rater bias in the assessment of social-emotional competence. *International Journal of Emotional Education*, *8*(2), 25.
Shapiro, V. B., Kim, B. K., Robitaille, J. L., & LeBuffe, P. A. (2017). Protective factor screening for prevention practice: Sensitivity and specificity of the DESSA-Mini. *School Psychology Quarterly*, *32*(4), 449. https://doi.org/10.1037/spq0000181
Shapiro, V. B., Lee, J., & Kim, B. K. E. (2022). TOOLBOX effects on K-2 student growth in social-emotional competence: Evidence from a quasi-experimental study. *Manuscript Submitted for Publication*.
Shulman, C. (2016). Social and emotional development in infant and early childhood mental health. In *Research and Practice in Infant and Early Childhood Mental Health* (pp. 23–42). Springer.
Simmons, C. A., Shapiro, V. B., Accomazzo, S., & Manthey, T. J. (2016). Strengths-based social work: A meta-theory to guide social work research and practice. In *N. Coady & P. Lehmann (Eds.), Theoretical Perspectives for Direct Social Work Practice* (3rd ed., pp. 131–154). Springer.
Skiba, R. J., Horner, R. H., Chung, C.-G., Rausch, M. K., May, S. L., & Tobin, T. (2011). Race is not neutral: A national investigation of African American and Latino disproportionality in school discipline. *School Psychology Review*, *40*(1), 85–107.
Skiba, R. J., Michael, R. S., Nardo, A. C., & Peterson, R. L. (2002). The color of discipline: Sources of racial and gender disproportionality in school punishment. *The Urban Review*, *34*(4), 317–342.

Skipper, Y., & Fox, C. (2021). Boys will be boys: Young people's perceptions and experiences of gender within education. *Pastoral Care in Education*, 1–19.
Sklad, M., Diekstra, R., Ritter, M. D., Ben, J., & Gravesteijn, C. (2012). Effectiveness of school-based universal social, emotional, and behavioral programs: Do they enhance students' development in the area of skill, behavior, and adjustment? *Psychology in the Schools*, *49*(9), 892–909.
Sørlie, M.-A., Hagen, K. A., & Nordahl, K. B. (2020). Development of social skills during middle childhood: Growth trajectories and school-related predictors. *International Journal of School & Educational Psychology*, 1–19.
Spinrad, T. L., & Eisenberg, N. (2009). Empathy, prosocial behavior, and positive development in schools. In *R. Gilman, E. S. Huebner, & M. J. Furlong (Eds.) Handbook of Positive Psychology in Schools*. Routledge.
Sue, D. W. (2010). *Microaggressions and Marginality: Manifestation, Dynamics, and Impact*. John Wiley & Sons.
Taylor, R. D., Oberle, E., Durlak, J. A., & Weissberg, R. P. (2017). Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. *Child Development*, *88*(4), 1156–1171. https://doi.org/10.1111/cdev.12864
Ura, S. K., & d'Abreu, A. (2022). Racial bias, social–emotional competence, and teachers' evaluation of student behavior. *Children & Schools*, *44*(1), 17–26.
U.S. Department of Education, Office of Elementary and Secondary Education. (2016). *Non-Regulatory Guidance: Student Support and Academic Enrichment Grants*. https://www2.ed.gov/policy/elsec/leg/essa/essassaegrantguid10212016.pdf
Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–70.
Verlenden, J., Naser, S., & Brown, J. (2021). Steps in the implementation of universal screening for behavioral and emotional risk to support multi-tiered systems of support: Two case studies. *Journal of Applied School Psychology*, *37*(1), 69–107.
Wang, R., & Ware, J. H. (2013). Detecting moderator effects using subgroup analyses. *Prevention Science*, *14*(2), 111–120.
Weissberg, R. P., Durlak, J. A., Domitrovich, C. E., & Gullotta, T. P. (2015). Social and emotional learning: Past, present, and future. In *J. A., Durlak, C. E. Domitrovich, R. P. Weissberg, & T. P. Gullotta (Eds.), Handbook of Social and Emotional Learning: Research and Practice* (pp. 3–19). Guilford Publications.
West, M. R., Pier, L., Fricke, H., Hough, H., Loeb, S., Meyer, R. H., & Rice, A. B. (2020). Trends in student social-emotional learning: Evidence from the first large-scale panel student survey. *Educational Evaluation and Policy Analysis*, *42*(2), 279–303.
Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*, *4*(1), 10–18.
Wilson, S. J., & Lipsey, M. W. (2007). School-based interventions for aggressive and disruptive behavior: Update of a meta-analysis. *American Journal of Preventive*

Medicine, *33*(2), S130–S143. https://doi.org/10.1016/j.amepre.2007.04.011
Wittmer, D., Doll, B., & Strain, P. (1996). Social and emotional development in early childhood: The identification of competence and disabilities. *Journal of Early Intervention*, *20*(4), 299–318.
World Health Organization. (1994). *Life skills education for children and adolescents in schools. Pt. 3, Training workshops for the development and implementation of life skills programmes* (WHO-MNH-PSF-93.7A.Rev.2). World Health Organization.
Yang, Y., Luo, Y., & Zhang, Q. (2020). A cautionary note on identification and scaling issues in second-order latent growth models. *Structural Equation Modeling: A Multidisciplinary Journal*, 1–12.
Yates, T., Ostrosky, M. M., Cheatham, G. A., Fettig, A., Shaffer, L., & Santos, R. M. (2008). *Research Synthesis on Screening and Assessing Social-Emotional Competence*. The Center on the Social and Emotional Foundations for Early Learning.
Zahn-Waxler, C., & Smith, K. D. (1992). The development of prosocial behavior. In *The development of prosocial behavior* (Van Hasselt V. B. & Hersen M. (Eds.) Handbook of Social Development: A Lifespan Perspective, pp. 229–256). Springer.
Zarbatany, L., Hartmann, D. P., Gelfand, D. M., & Vinciguerra, P. (1985). Gender differences in altruistic reputation: Are they artifactual? *Developmental Psychology*, *21*(1), 97.
Zhang, D., & Katsiyannis, A. (2002). Minority representation in special education: A persistent challenge. *Remedial and Special Education*, *23*(3), 180–187.
Zins, J. E., Bloodworth, M. R., Weissberg, R. P., & Walberg, H. J. (2004). The scientific base linking social and emotional learning to school success. In *J. E. Zins, R. P. Weissberg, M. C. Wang, & H. J. Walberg (Eds.) Building Academic Success On Social And Emotional Learning: What Does The Research Say?* (pp. 3–22). Teachers College Press.
Zoric, T. (2014). Poverty and class bias in schools. In *E. L. Brown, P. C. Gorski, & G. Lazaridis (Eds.) Poverty, Class, and Schooling: Global Perspectives on Economic Justice and Educational Equity* (pp. 347–364). IAP.