

What's in a Rating?  
Teacher Evaluations of Students' Social Emotional Competency:  
A Mixed Methods Study

Michelle Alissandra Proulx

A dissertation submitted in partial fulfillment of the  
requirements for the degree of

Doctor of Philosophy

University of Washington

2017

Reading Committee:

Leslie Herrenkohl, Chair

Todd Herrenkohl

Clayton Cook

Liliana Lengua

Program Authorized to Offer Degree:

College of Education

©Copyright 2017

Michelle Alissandra Proulx

## **Abstract**

### What's in a Rating? Teacher Evaluations of Students' Social Emotional Competency: A Mixed Methods Study

Michelle Alissandra Proulx

Chairperson of the Supervisory Committee  
Professor Leslie Herrenkohl  
Department of Learning Sciences and Human Development

The impact of social and emotional skills on educational, career, and life outcomes is well supported in the research literature. Schools are increasingly implementing universal curricula to teach social and emotional skills. Current school accountability requirements place pressure on school administrators and policy makers to provide evidence that social and emotional learning programs are effective in helping students to develop the non-academic skills associated with school, career, and life success. Yet, the body of research investigating universal social and emotional competency measures as part of school practice is limited. The literature offers few studies that address the ways in which teachers are taking up the practice of assessing social and emotional skills and competencies. The current sequential explanatory mixed methods study investigated teacher ratings ( $N = 173$ ) on the DESSA-Mini (Naglieri, LeBuffe, & Shapiro, 2009) for students in grades 3-5 ( $N = 4005$ ). The teachers worked in 17 elementary schools in a

minority-majority district in the Pacific Northwest. Hierarchical linear models were developed to evaluate the relationships of child, classroom, and school variables and teacher ratings. When controlling for classroom and school membership, a child's gender, age relative to classmates, race/ethnic heritage, and eligibility for specific support services significantly predicted teacher ratings. Predictor effects were moderated by school context. Following the quantitative investigation, 15 teachers were purposively sampled to engage in semi-structured cognitive interviews. Think-aloud, verbal probing, and expansive probing techniques were used to explore the ways in which teachers' thoughts, beliefs, biases, and experiences influenced their ratings of students' social and emotional competencies. Three response prototypes emerged from participants' think-aloud ratings and response styles were associated with informants' ratings of student behaviors. Significant differences were found in participants' interpretations of rating scale prompts and the response choice categories. Furthermore, teachers expressed concern about applying personal standards of behavior to diverse groups of students. The current study highlights the need for more rigorous pre-testing of assessment instruments to improved consistency in teachers' interpretations of the measure. Additional training is also needed to calibrate the ways in which teachers engage in the practice of completing rating scale measures. The current study calls into question the use of teacher ratings for cross classroom/ school

comparisons, and as measures of student growth when ratings are completed by different informants.

## Table of Contents

List of Figures.....	ii
List of Tables.....	iii
Chapter 1: Introduction.....	1
Chapter 2: Measurement of Student Behavior and Social and Emotional Skills.....	17
Chapter 3: Conceptual Model and Theoretical Framing.....	41
Chapter 4: Method.....	56
Chapter 5: Results .....	80
Chapter 6: Discussion .....	155
Chapter 7: Conclusion, Limitations, Implications and Future Research.....	168
References.....	176

## List of Figures

<i>Number</i>		<i>Page</i>
Figure 1.1.	<i>CASEL social and emotional learning competencies</i> .....	3
Figure 3.1.	<i>Cognitive response to rating scale measure</i> .....	50
Figure 3.2.	<i>Factors influencing teacher ratings</i> .....	54
Figure 4.1.	<i>Coding process (adapted from Castro et al. model, 2010)</i> .....	74
Figure 5.1.	<i>Cognitive response types and associated response qualities</i> .....	124

## List of Tables

<i>Number</i>		<i>Page</i>
Table 1.1.	<i>CASEL competencies and related skills</i> .....	22
Table 3.1.	<i>Components of the response process</i> .....	49
Table 4.1.	<i>DESSA scales and sample items</i> .....	60
Table 4.2.	<i>Summarization of independent variables</i> .....	64
Table 4.3.	<i>School SES, classroom means, and informant descriptors for interviewees</i> .....	70
Table 4.4.	<i>Congruence rubric for evaluating teacher definitions and ratings</i> .....	78
Table 5.1.	<i>Inter-item correlations among DESSA-Mini Items</i> .....	80
Table 5.2.	<i>Descriptive statistics for individual DESSA-Mini items</i> .....	80
Table 5.3.	<i>Response frequencies for individual DESSA-Mini items</i> .....	81
Table 5.4.	<i>Variance within-student ratings</i> .....	82
Table 5.5.	<i>SET Means and standard deviations for child factors</i> .....	83
Table 5.6.	<i>SET Means and standard deviations for child factors at Title 1 Schools</i> .....	84
Table 5.7.	<i>SET Means and standard deviations for child factors at Non-Title 1 Schools</i> .....	85
Table 5.8.	<i>Correlations among student level variables</i> .....	86
Table 5.9.	<i>Correlations among classroom level variables</i> .....	86
Table 5.10.	<i>Correlations among school level variables</i> .....	87
Table 5.11.	<i>Descriptive statistics for variables at each level of model</i> .....	88
Table 5.12.	<i>Descriptive statistics for variables at Title 1 schools</i> .....	89
Table 5.13.	<i>Descriptive statistics for variables at Non-Title 1 schools</i> .....	90
Table 5.14.	<i>Partitioning of variance for total sample</i> .....	90
Table 5.15.	<i>Partitioning of variance for Title 1 schools</i> .....	91
Table 5.16.	<i>Partitioning of variance for Non-Title 1 schools</i> .....	92
Table 5.17.	<i>HLM level 1base model by social emotional total for full sample</i> .....	93
Table 5.18.	<i>HLM level 1base model by social emotional total with interactions</i> .....	95
Table 5.19.	<i>HLM level 1base model by social emotional total for Title 1</i> .....	96
Table 5.20.	<i>Title and Non-Title level 1 variable group variance contributions</i> .....	97
Table 5.21.	<i>Effects of School Type on Child Level Predictors of SET Total</i> .....	99
Table 5.22.	<i>HLM level 2 base model by social emotional total</i> .....	100
Table 5.23.	<i>HLM level 2base model by social emotional total for Title 1 and Non-Title 1</i> .....	101
Table 5.24.	<i>HLM level 3 base model by social emotional total</i> .....	101
Table 5.25.	<i>Coefficients from HLM analyses of SET total score</i> .....	103
Table 5.26.	<i>Comparison of participating classrooms to full sample</i> .....	106
Table 5.27.	<i>Comparison of Non-Title 1 participating classrooms to full sample</i> .....	107
Table 5.28.	<i>Comparison of Title 1 participating classrooms to full sample</i> .....	108
Table 5.29.	<i>Teacher demographics and descriptive statistics for classroom ratings</i> .....	109
Table 5.30.	<i>Comparison of participating teachers to full sample</i> .....	110
Table 5.31.	<i>Child characteristics of students rated in strength range</i> .....	111
Table 5.32.	<i>Child characteristics of students rated in typical range</i> .....	111
Table 5.33.	<i>Child characteristics of students rated in need range</i> .....	112
Table 5.34.	<i>Teacher demographics and statistics for classroom ratings by response type</i> .....	119

## **Acknowledgements**

This project is the culmination of over twenty years of study on how best to address the social and emotional needs of children within the school setting. Many teachers, friends, and colleagues have joined with me to think about how schools can meet the social and emotional needs of students and have offered support and encouragement when the work felt too big. I would like to acknowledge my dissertation committee members Leslie Herrenkohl, Todd Herrenkohl, Clayton Cook, and Liliana Lengua their sustained their interest in my work and support over the course of the project. I would also like to express my gratitude to faculty members who shaped my early thinking about human development, assessment, and emotional well-being, Sam Wineburg, Cathy Taylor, Ginger Berninger, and Gene Gallagher.

This project would not have been possible without the statistical guidance and support of Gabby Gorsky, William Brown, and Colin Beam from The Center for Social Science Computation and Research at the University of Washington. I would also like to recognize my colleagues who have worked to advance teacher understanding that the social and emotional well-being of students is the foundation for academic achievement and healthy school communities. Your efforts have improved services for children during the school day and have enhanced the skills of adults who model behaviors for students every day. Thank you Judy Buckmaster, Naomi Calvo, Stephanie Wright, and Randi Peterson for your ongoing advocacy and commitment to the cause. I would also like to thank the teachers who participated in teacher interviews. Your honesty, commitment to children, and desire to ‘get it right’ give me great hope for the future of this work.

Finally, to my family and friends who have held an unwavering belief that I could complete this work and cheered me on when I needed it most. Michael Quirk, Rita Quirk, Peggy

Quirk, Maureen Rielly, Edward Rielly, Patricia Logan, Patsy Naughton, Leihua Edstrom, Cindy Benge, Becky Street, Mary-Jo Stewart, Sarah Geren-Ziegler, Stacy Sikora, Phyllis Goodman, Linda Brooks, Maureen Toll, and Erik Kaplan your encouragement has meant the world to me.

## **Dedication**

I dedicate this dissertation to my husband Greg Proulx, who has selflessly and steadfastly supported me and my dreams. Your enduring love and encouragement has made it all possible. The discussions on our long walks can now return to topics of daily life!

To our children, Michael, Emily, and Katherine, you have been my most important teachers. Each of you has trusted me with your heart. It has been an honor to walk alongside you during difficult times and celebrate with you as you achieved your goals. The greatest lessons of patience, perseverance, and letting go, however, were learned on ordinary days as we shared our lives, working, playing, arguing, and loving each other.

I would also like to recognize the children I have worked with over the years who have inspired me with their spirit, resilience, and forgiving hearts. This work is for you!

## **Chapter 1: Introduction**

Educators are increasingly focused on students' non-academic skills and the impact these capabilities have on academic achievement, classroom climate, and personal well-being. Historically, teachers addressed non-academic behaviors, along with character traits, through classroom rules, civics curricula, character programs, and/or bullying prevention lessons (Elias et al., 2015; Lockwood, 1993). Early attempts to develop non-academic skills were often in response to behaviors exhibited by individual students that compromised a positive classroom climate, or interfered with learning (e.g. physical aggression, bullying). Schools reacted to these behaviors by providing targeted interventions to identified children. Typically, these efforts focused on a narrow set of skills (e.g. emotion regulation, conflict resolution) taught by school counselors outside the classroom environment. The underlying theories guiding skill development in these targeted interventions was often not intentionally aligned with existing initiatives within the school setting. Moreover, the language and cuing, introduced to support the development and performance of new skills, was seldom reinforced by the classroom teachers. Traditionally, interventions for identified students were delivered in short weekly sessions, targeted to a specific set of skills, and lasting only a few months (Domitrovich et al., 2009; Domitrovich & Greenberg, 2000; Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Greenberg, Domitrovich, Graczyk, & Zins, 2005; Stephan, Sugai, Lever, & Connors, 2015; Walker & Horner, 1996). The growing social and emotional learning (SEL) movement, in contrast, advocates for systemic, coordinated, and sustained delivery models that provide universal and targeted evidence-based skill instruction to all students. SEL models are structured to simultaneously prevent negative outcomes, treat existing social and/or emotional difficulties, and promote well-being.

Social and emotional learning (SEL) emerged as a field following an interdisciplinary conference of researchers, child advocates, and educators who recognized the benefits of a unified approach and research agenda for addressing the needs of the whole child (Durlak, 2015). The Collaborative for Academic Social and Emotional Learning (CASEL) was established and the generic term SEL is now used to describe a range of cognitive, affective, and behavioral competencies believed to be pivotal to students' social and emotional competence, academic performance, and life outcomes (Collaborative for Academic, Social, and Emotional Learning (CASEL), 2013; Elias, 1997; Durlak, 2015; Weissberg & Cascarino, 2014). In 1994, CASEL introduced a conceptual framework of essential competencies that reflect the research contributions that various research traditions (e.g. prevention science, positive youth development, affective psychology, and clinical psychology) have made to the emerging field. The framework focuses on 3 broad interrelated competency areas: (a) emotional competence; (b) social competence; and (c) responsible decision making.

More specifically, CASEL conceptualizes these competency areas in five multi-dimensional domains that address discrete, but related skills (Collaborative for Academic, Social, and Emotional Learning (CASEL), 2013; Kendziora, Weissberg, & Dunesbury, 2011). These domains include: 1) self-awareness; 2) self-management; 3) social awareness; 4) relationship skills; and 5) responsible decision making (See Figure 1.1).

Social & Emotional Learning Core Competencies



**Figure 1.1.** CASEL social and emotional learning competencies

The intrapersonal skills comprising the self-awareness (e.g. assessing one’s emotion) and self-management (e.g. regulating, or expressing ones’ emotion) domains are rooted in the highly-researched construct of emotional competence and the subfield of emotion regulation. Social awareness (e.g. perspective taking) and relationship skills (e.g. social problem solving, cooperation) domains draw upon skills included within the broader field of social competence which includes social problem solving and relationship skills. Responsible decision making is perhaps the competency area receiving the least attention, and includes the skills associated with making decisions in accordance with ethical standards, safety, and social norms (Collaborative for Academic, Social, and Emotional Learning (CASEL), 2013).

The field of social and emotional learning is built upon prior work and research findings from related disciplines. Accordingly, this introduction cuts across the fields of psychology, prevention science, and positive youth development (and others) to highlight research documenting the impact of non-academic skills on the achievement and well-being of students, the health and culture of school communities, and on more distal outcomes associated with career and life success. The terminology and conceptualization of specific skills within each of these traditions is not wholly shared; however, the academic literature emerging from each of

these research communities has contributed to our understanding of how social and emotional skills are associated with important school and life outcomes. A comprehensive review of the literatures from each of the fields contributing to the newly formed social and emotional learning discipline is outside the scope of this project. The included review, instead, is intended to be illustrative, and is presented to highlight the work that SEL scholars, policy makers, and educators draw upon to support the inclusion of explicit social and emotional skill instruction during the school day.

### **Social and Emotional Competencies**

**Emotional competence.** Emotional competence is a term that describes capabilities associated with the skills individuals use as they experience and act upon emotion. These skills include the awareness of personal emotions, the ability to name emotions, an understanding of social norms governing the expression of emotion, and the ability to regulate the intensity and duration of emotional states (Eisenberg & Spinrad, 200). Emotional competence in the school setting is typically associated with low levels of emotional negativity and high levels of emotion regulation. Emotionally competent students are skillfully able to modulate their emotions to maximally utilize cognitive resources for learning (Eysenck, Derakshan, Santos, & Calvo, 2007; Eysenck, 2013). While personality and temperament research has highlighted the biological basis of dispositional emotionality and reactivity, ample scientific evidence suggests the skills associated with managing internal emotional states and behavioral responses can be learned (Goleman, 1995; Humphrey, Curran, Morris, Farrell, & Woods, 2007; Rice, Levine, & Pizarro, 2007; Salovey & Mayer, 2004; Weare & Gray, 2003).

Arguably, the skills associated with emotional competency warrant the attention of educators, child advocates, and professionals working in the field of youth development.

Research conducted within the field of emotion regulation demonstrates that students who effectively manage their emotions have an increased ability to selectively apply their mental capacities to learning (Blair, 2002; Duncan et al., 2007; Garner, Mahatmya, Brown, & Vesely, 2014; Hinshaw, 1992; Rice et al., 2007). In contrast, students reacting to, or overly focused on, their emotions utilize valuable cognitive resources to attend to emotions which may detract from learning and threaten academic achievement (Blair, 2002; Blankstein, Toner, & Flett, 1989; Eysenck et al., 2007; Eysenck, 2013; Garner, 2010). The loss of instructional time is especially costly for students who are unable to manage their emotional state through effortful control (see Kim, Nordling, Yoon, Boldt, & Kochanska, 2013).

Research has documented the negative impact of poor emotion regulation skills on students' academic achievement. Low emotional competence compromises instructional time (Greenwood, 1991), interferes with test performance (Eysenck et al., 2007), impedes the recall of instructional content (Garner, 2010; Rice et al., 2007), lowers task-persistence (see Elliott, Malecki, & Demaray, 2001; Schutz & Davis, 2000), and decreases school engagement (Masten & Cicchetti, 2010). Additionally, students with less developed emotional skills are perceived by their teachers as more difficult to teach and less capable than their emotionally competent peers (Eisenberg, Sadovsky, & Spinrad, 2005; Garner, 2010; Keogh & Burstein, 1988; Martin, Diane, Gaddis, & Moseley, 1988).

Emotional competence also plays a role in shaping students' interpersonal relationships. Students with strong emotion regulation skills who exhibit low levels of negative emotion are likely to enjoy positive social interactions with others (Caspi, 1995; Frey, Nolen, Van Schoiack Edstrom, & Hirschstein, 2005; Jones, Brown, & Lawrence Aber, 2011). Students adept at managing their emotional states are also less likely to exhibit aggressive behaviors directed

toward others and are less likely to become targets of bullying behavior (Garner, 2010; Hanish et al., 2004). When encountering unpleasant experiences, emotionally competent students are more apt to restructure unhelpful, or hostile thoughts, by using cognitive abilities to re-appraise experiences in a manner that supports positive views of others (Elliott et al., 2001; Spivack, Platt, & Shure, 1976). Students who are able to understand and manage emotions are more likely to exhibit prosocial empathy behaviors (Denham, 1986; Garner, 2010; Schonert-Reichl, Smith, Zaidman-Zait, & Hertzman, 2012), engage in more positive interactions with others (Brackett, Mayer, & Warner, 2004; see Eisenberg, Fabes, Guthrie & Reiser, 2000 for review; Lopes, Salovey, & Straus, 2003), have more friends (Ladd & Birch, 1997), and enjoy greater peer popularity (Garner, 2010).

The skills associated with emotional competency have been shown to have a cascading effect on individual adjustment and positive well-being over the life span (Eisenberg, Shepard, Fabes, Murphy, & Guthrie, 1998; Masten & Cicchetti, 2010). Early emotion regulation skills are predictive of later mood management and personal well-being (Cicchetti, Ganiban, & Barnett, 1991). Numerous research studies clearly connect emotion regulation capabilities with mental health outcomes (Eisenberg et al., 1995, 2001, 2000, 1998). Children with high emotional negativity and low emotional control are prone to exhibit under-controlled and aggressive behaviors (e.g. hitting, kicking, yelling, etc.) which may contribute to conduct disorders, violence and illegal behaviors. Children who over-control their emotions, in contrast, are prone to internalizing behaviors and can be withdrawn, displaying a restricted range of emotion (Block & Block, J., 1980), placing them at risk for mood disorders including anxiety and depression.

**Social competence.** Social competence is a term used to describe a second set of capabilities associated with the display of ‘socially acceptable learned behaviors’ (Malecki &

Elliot, 2002). Interrelated with emotional competency, social competence includes the ability to collaborate, cooperate, form and maintain relationships, take the perspective of others, and resolve interpersonal conflicts (Collaborative for Academic, Social, and Emotional Learning (CASEL), 2013; Elias, Ferrito, & Mocerri, 2016). These skills are increasingly called upon in current school initiatives (e.g. STEM, project based learning, NGSS) which require students to work together to analyze and solve unfamiliar problems that may have numerous solutions (Bellanca, 2011; National Research Council, 2012).

Social competency skills are positively correlated with academic engagement, academic achievement and several other educational outcomes. Students who exhibit prosocial skills are more apt to sustain involvement in interpersonal exchanges in which ideas are considered (Elliot et. al. 2001). The increased time and engagement in questioning, discussing, and debating ideas provides socially competent students with more opportunities to enhance their knowledge and skills (see Bandura, 1977; see Vygotskii & Cole, 1978). Teachers' ratings of students' social skills are positively correlated with their grade point averages (Malecki & Elliot, 2002). An association between a teachers' ratings of students' social skills and their performances on nationally standardized measures of academic achievement--albeit weaker—are also observed (Wentzel, 1993). These skills also predict future academic performance (Caprara, Barbaranelli, Pastorelli, Bandura, & Zimbardo, 2000). Carrera and colleagues (2002) found social skills in early elementary school are better predictors of academic achievement in middle school than early elementary reading scores.

Additionally, socially competent students are more likely to be viewed favorably by teachers and peers and these positive connections support academic engagement and a commitment to the school community (Caprara et al., 2000; Hughes, Gleason, & Zhang, 2005;

Oberle, Schonert-Reichl, Hertzman, & Zumbo, 2014; Saft & Pianta, 2001). Students who enjoy positive high quality relationships with teachers are more likely to succeed academically than students who experience conflicted or strained relationships (Darling-Hammond & Friedlaender, 2008; Friedlaender, Burns, Lewis-Charp, Cook-Harvey, & Darling-Hammond, 2014; Jones et al., 2011). Students who are unable to form positive relationships with others often struggle to meet the expectations of the school environment (Blair, 2002; Cooper, Moore, Powers, Cleveland, & Greenberg, 2014; Denham, 2006). These students more frequently receive negative feedback from members of the school community and are more often referred for disciplinary action which may result in time away from the classroom and missed instructional time. Individual student difficulties with social competence can also interfere with the academic achievement of classmates, as disruptive behaviors enacted by one student have the potential to upset the entire classroom environment and compromise instructional time (see Dishion, McCord, & Poulin, 1999; Jones et al., 2011; Kellam, Ling, Merisca, Brown, & Ialongo, 1998; Thomas, Bierman, & Powers, 2011).

In addition to the educational costs of poor interpersonal skills which include academic underachievement, school disengagement, disrupted classroom communities, and school dropout (Jones et al., 2011; Oberle et. al, 2008), social skill deficits also contribute to negative life outcomes over the course of development. Individuals with poor social skills are more likely to be alienated from their communities and are at an increased risk for unsafe sexual behaviors, underemployment, involvement with violence and/or crime (Catalano, Hawkins, Berglund, Pollard, & Arthur, 2002; Domitrovich et al., 2009; Guerra & Bradshaw, 2008; Hawkins, Catalano, Kosterman, Abbott, & Hill, 1999; Heckman, Stixrud, & Urzua, 2006) and significant

mental health problems (Coie et al., 1993; S. A. Denham, Wyatt, Bassett, Echeverria, & Knox, 2009; Elliot & Gresham, 1987; Greenberg, Domitrovich, & Bumbarger, 2001).

**Responsible decision making.** Responsible decision making, the third broad competency area in the CASEL framework, is defined as “the ability to make constructive and respectful choices about personal behavior and social interactions based on consideration of ethical standards, safety concerns, social norms...” (<http://www.casel.org>). In practice, responsible decision making is often conceptualized as the ability to leverage skills included within the social and emotional competency domains to make decisions in accordance with personal values and localized social norms. In the school setting, students perceived to be making responsible decisions make personal choices to follow classroom rules, play fairly with others, resolve conflicts constructively, and set and academic and personal goals (Durlak, 2015).

### **Need for Social and Emotional Skills in Schools**

Provided with ample evidence documenting the impacts of emotional competency, social competency, and responsible decision making on student outcomes in school, work, and life, there is a growing urgency amongst educators, child advocates, and the mental health community to actively address the social and emotional development of children during the school day. The need is great. According to the U.S surgeon general, between 13% and 20% of American children exhibit symptoms associated with a mental health disorder (United States et al., 2001) and the prevalence of these disorders is growing (“Mental Health Surveillance Among Children — United States, 2005–2011,” 2013). In 2010, mood disorders were the most common primary diagnosis for children hospitalized in the U.S., and suicide became the leading cause of death for children between the ages of 12 and 17 years (“Mental Health Surveillance Among Children — United States, 2005–2011,” 2013).

The growing prevalence and severity of mental illness among America's children is also reflected in the increased use of mental health care services. Between the years of 2007-2010 the utilization of inpatient mental health, or substance abuse treatment services, grew by 24% ("Children's health care spending report: 2007-2010," 2012, "Mental Health Surveillance Among Children — United States, 2005–2011," 2013). These numbers likely represent only a fraction of those requiring care. In 2002, the surgeon general reported that merely 20% of children exhibiting symptoms associated with a mental health disorder receive care (Chafouleas, Volpe, Gresham, & Cook, 2010; President's New Freedom Commission on Mental Health, 2002) and even lower care access rates are observed amongst children of color (Lau, 2006; United States et al., 2001). Even amongst families who secure mental health care for their children, 40-60% of families discontinue services prematurely (Heathfield & Clark, 2004).

Schools are a logical setting for addressing the pressing social and emotional needs of children (President's New Freedom Commission on Mental Health, 2002). Most students attend school for six or more hours a day for most of the year. The schoolhouse provides consistent and sustained access to children over multiple years in which systemic, sequential, and explicit skill development can be delivered. Educational agencies and school districts are increasingly replacing multiple uncoordinated 'siloed' school interventions focusing on a single issue (e.g. substance use, absenteeism, bullying prevention, friendship skills) with comprehensive universal evidence based programs to achieve SEL targets (Domitrovich et al., 2009). Educators with daily access to children are uniquely positioned to explicitly teach students skills, model desired behaviors, and coach for skill performance in the context of ecologically valid classroom experiences. Furthermore, instruction and coaching in the child's natural environment preserves the complexity of real situational demands which are often missing in a more structured

therapeutic environment. In the school setting, students must attend to the verbal and non-verbal communication of others, manage personal emotions amidst distractions, and negotiate and respond to the varied expectations and demands of the environment. The demands students naturally encounter during a typical school day provide numerous experiences that require students to exercise social and emotional skills. School learning is an emotional experience. Learning can be exciting, boring, frustrating, encouraging, nerve-wracking, and surprising. Additionally, learning is often a social endeavor that presses students to communicate, collaborate, negotiate, and compromise with others. Schools provide an accessible and rich context for learning and practicing social and emotional skills (Hargreaves, 2000; Denzin, 2007).

**Teacher support for universal social and emotional skill instruction.** America's teachers support the inclusion of social and emotional instruction as part of the school day. In a nationally representative sample of pre-k through grade 12 teachers, 93% of teachers surveyed believe that SEL is very, or fairly, important for the 'in school student experience' (Bridgeland, Bruce, & Hariharan, 2013, p. 5). This finding holds true across school levels and type of school (e.g. public, private, religious). Of the educators surveyed, 95% of teachers believe that social and emotional skills can be taught, and 97% recognize that social and emotional instruction is important for all students, regardless of socioeconomic status. Teachers believe that an increased focus on social and emotional development as part of the school day has the potential to increase overall academic achievement (75%), improve standardized test scores (77%), improve attendance rates (80%), improve behavior (78%), and keep students on track to graduate (80%) (Bridgeland, Bruce, & Hariharan, 2013).

**Instruction benefits.** There is also mounting evidence that social and emotional skills can be taught (Goleman, 1995; Peter Salovey & Sluyter, 1997) and that enhanced skills

positively contribute to desirable outcomes in and out of school. The CASEL Guide endorsed 19 programs for elementary school student that meet CASEL standards for excellence based on well-designed curriculum, high quality training and implementation materials, and evidence to support program effectiveness (“CASEL Guide Online,” 2013). Empirical reviews of universal SEL programs demonstrate the positive impact these programs have on students’ skill development and numerous other outcomes including academic achievement, emotional well-being, classroom climate, and interpersonal relationships (Brackett, Rivers, Reyes, & Salovey, 2012; Committee for Children, 2013; Frey et al., 2005; Greenberg, Kusche, Cook, & Quamma, 1995; Sklad, Diekstra, Jehonathan, & Gravesteijn, 2012; Rivers, Brackett, Reyes, Elbertson, & Salovey, 2012; Schonert-Reichl et al., 2012; Schonfeld et al., 2015, 2015; Weare & Nind, 2011; Weissberg et al., 1981). In a landmark meta-analysis, Durlak and his colleagues (2011) consolidated findings generated from 213 studies conducted over 20 years to document the impact of SEL programs have on student outcomes. Compared to control group subjects, students receiving sequential, active, focused, and explicit SEL instruction exhibited improved social and emotional skills, more positive attitudes toward self and others, and more positive social behaviors. Treatment group students receiving SEL instruction also exhibited fewer conduct problems, and reported lower levels of emotional distress. Additionally, students taught social and emotional skills academically outperformed their control group peers.

### **Legislation and Policy Efforts to Support Social and Emotional Learning in Schools**

In response to burgeoning research documenting the impact social and emotional skills on school, work, and life outcomes, federal lawmakers have introduced legislation to support the delivery of evidence-based SEL programs in schools. In 2011, a bill to expand the Elementary and Secondary Education Act (ESEA) called for funding to provide professional development

and technical assistance to support the implementation and delivery SEL programs in K-12 schools (Academic Social and Emotional Learning Act, H.R. 2437). The bill proposed that teacher and principal training include instruction in the knowledge and skills necessary to support the development of student' social and emotional skills. Despite bipartisan support, the bill failed to advance beyond committee hearings. Subsequently, the Supporting Social and Emotional Learning bill was reintroduced in 2013, 2014, 2015, reflecting a continued commitment advocating for including SEL in the school context ("Academic, Social, and Emotional Learning Act of 2013" (2013 - H.R. 1875), "Supporting Emotional Learning Act" (2014 - H.R. 4509), "Supporting Social and Emotional Learning Act", 2015; Zaslow, Mackintosh, Mancoll, & Mandell, 2015). In 2015, Every Student Succeeds Act (ESSA) passed and includes provisions for an increased focus on students' non-academic skills. Most importantly, the act includes an expanded definition of student success which recognizes school engagement, effective communication skills, and skills that contribute to a positive school climate as important contributors to school success, reversing the constricted and purely academic notions of school success precipitated by the No Child Left Behind Act (2002). The new law also permits the use of non-academic indicators (e.g. school climate, teacher engagement) as school accountability measures to document school improvement ("Federal Policy and Legislation," "The Every Student Succeeds Act," 2015).

In addition to federal legislative efforts in support of school SEL efforts, at least three states (e.g. Illinois, Pennsylvania, Kansas) have developed comprehensive free-standing SEL standards for school age students (K-12). Articulated standards and developmental benchmarks clearly communicate social and emotional goals and learning targets to educators, families and community partners, and elucidate a shared vision for SEL efforts. The Illinois and

Pennsylvania Standards, enacted in 2004 and 2012, respectively, are fully aligned with the CASEL framework and identify target skills in each of the 5 CASEL domains targeting emotional competency, social competency, and responsible decision making. Kansas' standards expand upon the CASEL framework to include character development goals such as respect, fairness, kindness and honesty ("Kansas Social, Emotional, and Character Development Model Standards," 2012). Following the lead of these pioneer states committing to SEL efforts in their schools, at least eleven other states have introduced early learning and/ or elementary SEL standards to guide universal social and emotional instruction (Dunesbury, Weissberg, & Domitrovich, 2014).

### **Future Directions for SEL Programs**

Although the academic research literature strongly supports the delivery of universal SEL programs during the school day, information is only beginning to emerge about how these programs are functioning in practice. While over 500 program evaluations investigating the potential of SEL programs to contribute to positive student outcomes have been conducted under research controlled conditions, few studies illustrate how students are benefiting from SEL efforts in practice. In contrast to typical practice conditions, program evaluation studies typically provide professional development for teachers delivered by subject matter experts. Such training supports teachers to acquire the content knowledge and instructional practice experience to ensure quality instruction and adequate program support. Typically, efficacy studies also provide curriculum delivery oversight by research staff to assure the quality and dosage of instruction delivery. Furthermore, assessments to measure student skill development and evaluate program effectiveness are often conducted by research staff.

We are only beginning to learn about the uptake, implementation, and evaluation of SEL programs under real-world conditions. Little is known about how schools are balancing SEL goals with academic standards based initiatives such as Common Core State Standards (CCSS), and Next Generation Science Standards (NGSS), or how teachers, with little or no pre-service social and emotional theory training (Domitrovich et al., 2009; Greenberg, 2010; Jones & Bouffard, 2012), are responding to school, district, and state mandates to teach social and emotional skills. While the Missing Piece Survey (Bridgeland et al., 2013), suggests that teachers support the inclusion of SEL instruction during the school day, the quantitative studies that have dominated the field of SEL research do not adequately communicate how teachers are engaging in the practice of teaching social and emotional skills.

Perhaps no greater challenge faces the field of SEL than assessing whether SEL efforts are efficacious in producing the desired outcomes they target under ecologically valid conditions. Under real-world circumstances, are these programs improving students' skills? Are they contributing to fewer discipline referrals, enhanced classroom climates, and higher graduation rates? Educators, researchers, and policy makers are wrestling questions about how, what, and when to assess to answer these questions. Do we evaluate individual student skills? Do we assess the impact of skills on classroom or the school climate? State departments of education, school districts, and educational agencies are eager to ascertain whether the programs and practices they fund are achieving the desired outcomes to justify the use of instructional class time. Researchers within the field are also questioning whether particular groups of students experience differential program benefits based on academic ability, baseline social and emotional competency, or ethnic/ racial group membership (Bloom & Michalopoulos, 2011; Rimm-Kaufman et al., 2014). To answer these questions, we need to learn more about the evaluation of

students' social and emotional skills and factors that influence the selection of school assessment practices.

## **Chapter 2: Measurement of Student Behavior and Social and Emotional Competency**

In this era of standards based education, school administrators and policy makers are seeking data to provide evidence that students are adequately equipped with the social and emotional skills needed to succeed in school, in work, and in life. Increasingly, academic institutions are pressed to assess whether children can perform the non-academic skills that have proven to be vitally important to school and life outcomes. In addition to screening, monitoring, and measuring students' skills, school administrators are also looking for ways to evaluate the effectiveness of universal SEL programs.

Historically, schools have played a limited role in formally evaluating students' social and emotional skills. Although teachers have provided qualitative observations about students' behavior, citizenship, effort, or ability to work with others to families, their reports have generally reflected individual educators' assessments of a student without a guiding methodology for observing, understanding, or evaluating student behavior (Elias et al., 2016). Teacher evaluations of students' behaviors, skills, and competencies were based on informal and non-standardized approaches to understanding student behavior.

As report cards and student records have become electronic, software with drop-down comment menus and standardized indicators of student behaviors have replaced teachers' hand written comments. In many schools, teachers are now required to assign numerical ratings (e.g. 1-4) or categorical designations (e.g. 'needs improvement,' 'satisfactory,' 'emerging') to describe non-academic behaviors, skills, and competencies selected by district leaders and/or report card committees. The indicators are often a mish-mash of study skills, effort, and social skills that fail to align with a theoretical developmental framework or specific SEL goals (Elias et al., 2016; McKown, 2015). In many instances, educators are expected to assign scores, or

ratings, with little information about how the behaviors are defined, and without ‘look-fors’ that signal behavioral competency at a developmental level. We know little about teachers’ thought processes when evaluating student behavior and lack information about the knowledge, beliefs, training, and experiences educators draw upon to evaluate students’ social and emotional competencies.

Teachers receive limited--if any--pre-service training in assessing students’ social and emotional skills. Social and emotional learning as a content area is only recently appearing in pre-service teacher education programs. Coursework is typically limited to a single class, sometimes taught in the psychology department, and almost always untethered from applications to teachers’ classroom practices (Leibbrand & Watson, 2010; Schonert-Reichl, Hanson-Peterson, & Hymel, 2015). SEL experts describe the ‘current efforts to incorporate coursework in the developmental sciences into teacher education training...[as] woefully inadequate’ (Schonert-Reichl et al., 2015, p. 410). With little training or instruction in social and emotional development, teachers are left to on their own to acquire and apply social and emotional theoretical knowledge to their emerging SEL instructional and assessment practices.

Educators’ involvement in more formal assessments of students’ non-academic skills, behavior, and social-emotional competencies--until recently--has been in the context of special education evaluations, and limited to children already exhibiting behaviors interfering with school functioning (Levitt, Saka, Hunter Romanelli, & Hoagwood, 2007). In the course of special education evaluations, teachers are typically interviewed about a child’s behavior and asked to complete teacher rating scales to screen for behaviors suggestive of a social or emotional disability (Chafouleas et al., 2010; Elliott et al., 2001; Shapiro & Heick, 2004; Sheridan, Maughan, & Hungelmann, 1999). If a child’s behavior meets the designated

thresholds, the student may be found eligible for special education services under the Individuals with Disabilities Education Act (“Individuals with Disabilities Education Act (IDEA)”). In these targeted evaluations, teachers are asked to complete clinically-based measures that focus on behaviors associated with DSM-V mental health diagnoses (e.g. aggression, whining, stealing, social withdrawal).

As SEL programs take root in schools, there is a growing need to develop assessment strategies that do more than establish eligibility for special education services for a select group of students. Universal measures are needed to identify students at-risk, track skill development, and evaluate program impacts. Currently, there is not a single behavior, or skill that serves as global outcome measure for social and emotional competency (see Chafouleas et al., 2010). Indeed, substantive debate remains as to which skills, traits, or dispositions have the greatest impact on student outcomes and should be measured by school programs. The CASEL competencies serve as a foundation for many states’ standards, district programs, and school initiatives; however, considerable variability remains in how educational organizations situate these competencies in relation to other valued skills, programs, curricula, and initiatives supported by the organization. The California Office to Reform Education (CORE), for example, targets 4 social and emotional competencies as learning outcomes in partnering school districts: growth mindset (see Dweck, 2006), self-efficacy (Bandura, 1977a), self-management (Blair, 2002, 2010; Payton et al., 2000; Zins, Bloodworth, Weissberg, & Walberg, 2004), and social awareness (Payton et al., 2000; Zins et al., 2004). Only two of the CORE identified competencies (e.g. self-management, social awareness) are based on the CASEL framework. Kansas’ SEL state standards, in comparison, target each of the 5 CASEL competency domains and included an expanded focus on ethical behaviors in the state’s Social Emotional Character

Development standards (“Kansas Social, Emotional, and Character Development Model Standards,” 2012). The related, but distinct, goals adopted by each of these organizations, illustrate the diverse ways in which schools, districts, and youth organizations value and prioritize specific combinations of skills, traits, and dispositions in their programs.

The diversity found in SEL standards reflects the values, priorities, and goals of specific communities and policy makers; however, the lack of consistency and alignment both within and across organizational systems poses a challenge when selecting or creating feasible, efficient, and psychometrically sound universal assessment measures. In addition to the lack of agreement about which skills to measure (Denham et al., 2009; Denham, 2006; Duckworth & Yeager, 2015), questions remain as to how the skills included in a measurement tool are defined (Humphrey et al., 2007; Watson & Emery, 2010; Wigelsworth, Humphrey, Kalambouka, & Lendrum, 2010), the appropriate modality for assessing particular skills, factors influencing various measures (Epstein, March, Conners, & Jackson, 1998; Mashburn, Hamre, Downer, & Pianta, 2006; Pas & Bradshaw, 2014; Wolcott & Williford, 2015). More recently, researchers are also calling into question the validity and appropriateness of universal measures for evaluating culturally and linguistically diverse groups of students (Li, Marion, Perie, & Gong, 2010).

The review that follows focuses on research relevant to universal measures of students’ social and emotional competencies for evaluating and tracking individual skill development. The survey is limited to measurement topics associated with assessing the execution, or performance of skills, rather than the comprehension of skills (see Lipton & Nowicki, 2009) since SEL programs target the outcomes associated with an individual’s ability to perform social and emotional skills in real-world contexts. The ability to encode, interpret, and think about

social and emotional experiences may be a precursor to an individual's ability to exhibit skilled behavior, however; it is the capacity *to perform* these cognitive skills in pursuit of prosocial goals that is paramount.

### **Measurement Content**

The CASEL competencies are widely adopted and serve as a basis for SEL learning standards in many states and school districts; and yet, numerous school agencies expand SEL goals to include character traits, skills, or dispositions (e.g. grit, growth mindset, mindfulness) that lie outside the CASEL framework (e.g. Kansas SECD). Moreover, many evidence-based SEL programs—especially those introduced prior to the publication of CASEL framework—are grounded in theoretical traditions that pre-date CASEL (e.g. emotional intelligence, social learning theory). Accordingly, the lessons and practices included in these programs in some instances stem from theoretical traditions that do not specifically target CASEL competency goals. Hence, the connections between the CASEL competencies, district standards, and curricula can sometimes be inexact, making the selection, or development, of assessment measures challenging. A uniting thread, connecting learning standards to instructional goals, curricula, and classroom practices strengthens assessment quality (Taylor, 2008). When the ties between these elements are frayed, missing, or weak, assessment measures can be poorly aligned with articulated goals, compromising assessment quality and the validity of interpretations derived from assessment data.

CASEL advocates for a common conceptual framework to establish a shared set of learning targets to guide instruction, program delivery, and assessment in school programs and youth service agencies. As previously introduced, the multi-faceted CASEL competency areas include: 1) self-awareness; 2) self-management; 3) social awareness; 4) relationship skills; and 5)

responsible decision making. The operationalized definitions of these skills is shown in Table 1.1 (Collaborative for Academic, Social, and Emotional Learning (CASEL), 2013; Kendziora et al., 2011).

**Table 1.1.** *CASEL Competencies and Related Skills*

Competency	Skills comprising competency area
Self-awareness	Being able to accurately assess one’s feelings, interests, values, and strengths, maintaining a well-grounded sense of self-confidence
Self-management	Being able to regulate one’s emotions to handle stress, control impulses, and persevere in overcoming obstacles; setting and monitoring progress toward personal and academic goals; expressing emotions effectively
Social-awareness	Being able to take the perspective of and empathize with others; recognizing and appreciating individual and group similarities and differences; recognizing and using family, school, and community resources
Relationship skills	Being able to establish and maintain healthy and rewarding relationships based on cooperation; resisting inappropriate social pressure; preventing managing, and resolving interpersonal conflict; seeking help when needed
Responsible decision-making	Being able to make decisions based on consideration of reason, ethical standards, safety concerns, social norms, respect for self and others, and likely consequences of various actions; applying decision-making skills to academic and social situations; contributing to the well-being of one’s school and community.

These competency constructs emerged from the multiple research traditions contributing to the field of SEL, which has created a collision of assessment constructs and terminology.

Measurement researchers are using the same terms to refer to different concepts and using different terms to refer to comparable constructs (Humphrey et al., 2007; Perez, Petrides, &

Furnham, 2005; Watson & Emery, 2010; Wigelsworth et al., 2010). The term ‘emotional intelligence,’ for example, is used to reference both skill-based (see Salovey & Mayer, 2004) and trait-based (Bar-On & Parker, 2000) capabilities associated with the recognition and management of emotion (Van Rooy, Viswesvaran, & Pluta, 2005). Similarly, the term ‘social skills’ has been differentially operationalized according to at least three theoretical perspectives: peer acceptance definitions; behavioral definitions; and social validity definitions (Elliot & Gresham, 1987). The theoretical differences that underlie specific conceptualizations of terms pertaining to social or emotional competence are essential to how these terms are operationalized and measured, and these nuances must be clear to individuals evaluating behavior. Selecting a measure designed to evaluate a trait-based conceptualization of emotion management would be inappropriate for assessing a competency construct (e.g. self-management) which is defined as skill-based.

Numerous researchers have compiled compendia to consolidate social and emotional learning measures and related constructs (see Denham, Ji, & Hamre, 2010; Haggerty, Elgin, & Woolley, 2011; Ringwalt, 2008; Sosna & Mastergeorge, 2005; Strive Task Force, 2013). As previously noted, several of the existing measures are grounded in research traditions other than SEL (e.g. grit, self-efficacy, growth mindset), and accordingly evaluate constructs originating in those traditions. Although many of the behaviors, skills, and dispositions included in these measures (e.g. mood regulation) relate to the CASEL competencies (e.g. self-management), the constructs typically differ in content and scope from the targets conceptualized in CASEL framework. Additionally, most of the psychometrically validated measures currently available are deficit focused and screen for maladaptive behaviors (e.g. aggression, whining, impulsivity) rather than asset-focused social and emotional competencies.

Denham and her colleagues' compendium of preschool and elementary assessment measures (2010) is notable for the publication's focus on the 5 CASEL core competencies and related contextual factors (e.g. classroom climate). All measures included within the collection evaluate at least one CASEL competency. The editors identify 18 scales for use with elementary age students. Most measures (78%) included in the review assess two or fewer competency areas, and lack sufficient breadth for use in school based assessment programs prizing efficiency, cost, and feasibility. Only one measure, the Devereux Student Strength Assessment (DESSA) (LeBuffe & Naglieri, 2008) targets all 5 CASEL core competency skill areas.

### **Measurement Modality**

Numerous measurement modalities have been developed to evaluate children's social and emotional skills. Each modality offers strengths and limitations for use as a universal measure. Seven approaches are widely used in school-based research and examples of each modality are included in the Compendium of Preschool Through Elementary School Social-Emotional Learning and Associated Assessment Measures (Denham et al., 2010).

They include:

- Naturalistic observations (e.g. observing child in classroom environment)
- Structured observations or performance (e.g. structured conflict, strange situation)
- Interviews with students (re: behavior, beliefs, competence, scenario)
- Interviews with others (e.g. teachers, parent, peers)
- Socio-metric techniques (e.g. peer nominations)
- Self-report measures (e.g. questionnaires, surveys, rating scales)
- Rating scales completed by others (e.g. teachers, parent, peers)

In this review, I describe each modality, present research related, and discuss affordances and constraints of the method as a universal measure of students' social and emotional competency.

**Naturalistic observations.** Naturalistic observations, or direct assessments, evaluate student behavior as it occurs under ecologically valid conditions. Observers often use

standardized measures to evaluate the frequency of student behavior (e.g. initiating social contact with peer), the response from others (e.g. praise, peer engagement), and/or the duration of behavior (e.g. crying). Some observation tools require the observer to monitor student behavior at specific time intervals to support the evaluator's ability to identify associations between exhibited behavior and setting conditions at the time of the behavior (e.g. independent seat work, group project). Naturalistic observations enable the observer to hypothesize about functions of a child's behavior and identify the ways in which actions are reinforced or punished in the student's environment. Observations can also include the collection of qualitative information in the form of running records which can offer rich descriptions of a child's behaviors and his interactions with others in the environment.

Assessing student behavior using naturalistic observations offers several important benefits. Observations assess a child's ability to execute social and emotional skills while attending to contextual factors and negotiating the demands of the school environment. Thus, direct observations may be less biased than measurement modalities that require the reporter to summarize observations over extended periods (Carter, Briggs-Gowan, & Davis, 2004). They may also be less influenced by a rater's cumulative experiences and knowledge of a child's abilities in other areas of development (e.g. academic achievement) (Shapiro, DuPaul, Lutz, & Kern, 2011). A substantial research literature documents the low levels of correspondence frequently observed between direct observations and indirect measures of the same child (Hoge, 1983; Thomas et al., 2011; Wolcott & Williford, 2015). Future work is needed to better understand the low rates of agreement between observations and less direct measures of students' skills and competencies.

Naturalistic observations have significant drawbacks as a universal assessment modality. Observations require significant time to complete. They are expensive to administer and may be an impractical choice for universal assessment. Observations may be more suitable for identifying externally expressed behaviors such as verbally contributing to group discussions, or initiating social contact. Observation measures may be less sensitive to identifying internalizing behaviors such as a student's appraisal style, or social withdrawal behaviors which are frequently overlooked in the school environment, particularly among academically accomplished students (Cook et al., 2011; Thomas et al., 2011). Furthermore, collecting observational data in limited windows of time may fail to capture significant, but infrequent behaviors (e.g. physical aggression) that are important markers of student skill development (Wolcott & Williford, 2015).

**Structured observations.** Structured observations involve setting up a task or situation that prompts students to engage in particular behaviors of interest (Frey et al., 2005). Raters observe the structured situation to document for the presence, frequency, or quality of behaviors to provide a measure of students' skills. The structured nature of the observations controls the situational conditions, providing some level of standardization of context and task demands across observations. While structured observations may provide more standardized testing conditions than other modalities (Watson & Emery, 2010), such tasks may fail to reflect the range of skills that could be observed over extended periods of time and contexts. Additionally, structured observations typically focus on a narrow set of skills, rather than the breadth of skills targeted by the CASEL framework, and most state and district standards. Hence, structured observations may contribute valuable information about students' skills, but be inefficient for use as a universal measure of numerous skill targets.

**Student interviews.** Interviews with students can provide useful information about their comprehension of social and emotional skills and abilities to interact with adults. In conversation with a teacher or evaluator, children can be prompted to appraise social interactions and offer suggestions for solving interpersonal conflicts. Their explanations offer valuable information about students' understanding of social interactions and insight into the attributions they ascribe to peers during conflict. In addition, children can be asked questions about personal emotions and how they respond to stressful situations to assess children's comprehension of emotion and self-management skills. While interviews with children may have advantages, the information gathered during interviews with elementary age children may be limited by a child's self-awareness (Denham et al., 2009; Humphrey et al., 2007) and heavily influenced by a child's verbal reasoning abilities (Carroll & Steward, 1984; Wigelsworth et al., 2010). Strong correlations are frequently found between a child's verbal intelligence and measures of social competence obtained through student interviews. Students who are verbally adept are perceived to be more socially skilled than their less verbal peers. The universal use of student interviews in a school setting would be problematic because linguistically diverse students are likely to find interviews challenging as they may lack the specific vocabulary required to accurately describe their emotions or social problem solving skills. Finally, interviews are more appropriate for assessing a child's comprehension of skills (know skills) and are not as well suited for evaluating a student's utilization of skill (go skills) with peers under real world conditions (Kim et al., 2013; Metcalfe & Mischel, 1999). While self-reports may help to identify students with internalizing behaviors (Cook et al., 2011), low correlations are consistently found between self-report measures and independent evaluators' reports of students' behaviors and competencies (Achenbach, McConaughy, & Howell, 1987; Keefer, 2015; Renk & Phares, 2004).

**Interviews with others.** Structured, semi-structured, and open interviews with parents and teachers are costly approaches for gathering information about student behaviors as they require significant time to complete. The benefits of conducting interviews include: increasing understanding of student in multiple contexts, gathering information about the antecedents and consequences of specific behavior. Additionally, parents and teachers who have known a child over a long period may have insights regarding the etiology of behaviors and may have suggestions about how best to support the student. The qualitative information gathered during interviews fails to generate competency scores that can be used to evaluate change over time.

**Socio-metric approaches.** Socio-metric techniques evaluate the degree to which a child is liked or disliked by his peers. In contrast to peer rating scales, socio-metric measures do not involve students rating the frequency of specific behaviors or skills; instead, peers are asked to nominate students who meet a criteria or condition (e.g. popularity). Peer-nomination is an example of a socio-metric measurement approach frequently used in academic research involving elementary age students (see Peery, 1979). Children are asked to list the names of three peers (typically in their grade level or classroom) they like the most (LM) and create a separate list of three peers they like the least (LL). From these lists of most-liked and least-liked peers, the number of nominations for each child is tallied and added ( $LL + LM$ ) or subtracted ( $LM - LL$ ) to provide a measure of *social impact* or *social preference* score. The inclusion of nominating least-liked children is an important complement to the most-liked list as peer acceptance and rejection are only slightly negatively correlated (Coie, Dodge, & Coppotelli, 1982).

The peer nominations yield composite scores for social impact and social preference and the combination of a child's scores yield 5 social status typologies. Children designated as

‘popular’ have high scores in both social impact and social preference. Students with high social impact and low social preference are classified as ‘Rejected’ children. Children with low social impact and low social preference scores are described as ‘neglected’ and those with high social impact and a mixed social preference are considered ‘controversial’ (Coie et al., 1982; Gresham & Stuart, 1992; Peery, 1979). Students earning less extreme scores on the social impact and social preference dimensions fall within the ‘average’ classification category.

The behaviors of children typed as neglected, rejected, popular, or controversial have proven to be consistent across studies and age groups (Coie et al., 1982; Gresham & Stuart, 1992) with rejected students frequently exhibiting under-controlled externalizing behaviors and neglected children more likely displaying over-controlled withdrawal and internalizing behaviors such as anxiety or depression. Popular children are likely to have social skills that are valued within the setting, whereas students characterized as controversial frequently possess strong social skills that may be accompanied by externalizing behaviors sometimes used to overpower other students. Although socio-metric measures may identify children exhibiting internalizing or externalizing behaviors, they are ill-suited for evaluating the performance of specific skills or the development of those specific skills over time.

Socio-metric measures offer an efficient approach to evaluating a student’s level of peer acceptance within a community and may serve as a quick screener to identify children at risk for maladjustment; however, socio-metric measures present some ethical issues. Eliciting the names of socially rejected peers may also exacerbate marginalization of rejected or neglected children (Asher & Hymel, 1981). Additionally, children with a minority status in a particular setting are less likely to be included within the ‘popular’ group and are more likely receive nominations contributing to a ‘controversial’ social status classification (Coie et al., 1982). Race effects may

be more attributable to minority status within a context and be less influenced by skill differences inherent in a racial group.

**Self-report measures.** Self-report measures include student surveys and rating scales in which the child is asked to report on his own behavior, attitudes, or beliefs. Self-report measures are efficient and inexpensive assessments that provide information about a child's personal view of his behaviors. The use of self-rating scales for elementary-aged children presents significant challenges. Young children typically lack the literacy skills to read and provide a written response. They also lack self-awareness (Denham et al., 2009; Humphrey et al., 2007) and tend to make global evaluations based on recent events (Wigelsworth et al., 2010). A second-grader rating problem solving skills, for example, may base his self-report on recent recess experiences, rather than looking more holistically at his patterns of behavior over time.

Even among older elementary school students with more expansive world views, self-reports may be better suited to assessing students' efficacy for social and emotional behavior, the value students place on behaviors, or their attitudes about behaviors or skills. Research finds that individuals are poor reporters on their own abilities. Social bias and desirability contributes to individuals overestimating their reported abilities. In various studies evaluating the correspondence between an individual's reported abilities and his actual abilities, results revealed there is only a slight relationship between self-reports and direct measures. Brackett and Mayer (2003), for example, a small correlation ( $r = 0.21$ ) between students' self-reported emotional intelligence and their actual emotional intelligence on a direct measure of ability.

**Parent and teacher rating scales.** Rating scales are the most the most widely used approach for evaluating students' non-academic skills and behaviors in the school setting (Canivez, Watkins, & Schaefer, 2002; Denham et al., 2009; Denham, 2015; Elliot & Gresham,

1987; Elliott et al., 2001; Stinnett, Havey, & Oehler-Stinnett, 1994) as they offer an efficient, inexpensive, and unobtrusive method for assessing multiple skills, observed over a period of time, in a variety of situations, with a single measure (Canivez et al., 2002; Van Horn, Atkins-Burnett, Karlin, Ramey, & Snyder, 2007). Rating scales typically instruct informants to respond to a set of prompts that target behavioral indicators of a skill (e.g. self-management).

Respondents rate the observed frequency (e.g. never, rarely, occasionally, frequently, very frequently) of specific behaviors (e.g. does something nice for somebody) on a Likert-type scale, and these ratings yield composite scores in a broad behavior category (e.g. social skills) or competency area (e.g. relationship skills). Scale scores describe the level of functioning within a more narrow set of skills (e.g. empathy) or behaviors (Gresham & Elliot, 1990). The scores generated from rating scales are indirect measures of student behavior that rely on informants' memory and judgment of observed student behavior.

Both parent and teacher rating scales are widely used in research studies and in practice, and are highly trusted as measures of child behavior and competency. These ratings frequently have significant and far reaching consequence, including their use in making mental health diagnoses and determining placement in special education programs (Shapiro & Heick, 2004). Parent and teacher rating scales are also relied upon as outcome measures in psychotropic drug trials (Conners, 1969), SEL program evaluations (Durlak et al., 2011), and behavioral intervention studies (Epstein et al., 1998; Hoge, 1983).

While rating scales offer a promising approach for assessing students' social and emotional skills, the research literature frequently finds only small to moderate correspondence among informants rating the same child (Achenbach et al., 1987; Berg-Nielsen, Solheim, Belsky, & Wichstrom, 2011; Canivez et al., 2002; Edebrock, 1984, 1984; Gross, Fogg, Garvey, & Julion,

2004; Hinshaw, Han, Erhardt, & Huber, 1992; Konold, Walthall, & Pianta, 2004; Powell, & Kamphaus, 1994; Renk & Phares, 2004; Ruffalo & Elliott, 1997; Wolcott & Williford, 2015). A landmark meta-analysis (Achenbach et al., 1987) of inter-rater agreement revealed that among informants observing a child in the same context (e.g. classroom) and sharing a similar role (e.g. teacher, teaching assistant, or pairs of parents), inter-rater agreement of problem behaviors is only moderate ( $r = 0.60$ ). Among respondents with different roles (e.g. parent and teacher) observing a child in different contexts (e.g. parent and teacher), ratings are only slightly correlated ( $r = 0.28$ ). Rating differences between informants observing children in different contexts may reflect true variability in child behavior resulting from situation specific conditions (Konold et al., 2004). Higher cross-informant associations are observed for externalized problem behaviors (Berg-Nielsen et al., 2011; Powell, & Kamphaus, 1994). Renk and Phare (2004), completed a similar meta-analysis investigating cross-informant ratings of children's social competency. Moderate associations exist between parent and teacher ratings ( $r = 0.36 - 0.48$ ). Higher ratings are observed between peer and teacher ratings ( $r = 0.48$ ) and significantly lower association are observed between self-reports of social competency and other raters' evaluations of student competency ( $r = 0.21-0.30$ ).

### **Focus on Teacher Ratings**

Despite the persistent finding that cross-informant agreement is limited, rating scales continue to be amongst the most widely used methods for evaluating children's behavior (Edebrock, 1984; Elliot & Gresham, 1987). Rating scales are prized for their efficiency, low cost, and ability to evaluate numerous behaviors in a single measure (Denham, 2015; Denham et al., 2010; Duckworth & Yeager, 2015). Teacher measures are particularly promising as a source of measurement as educators can be asked to evaluate students at designated time intervals and

they have exposure to numerous children over extended periods of time in a variety of situations. Importantly, teachers are also fluent in English which eliminates the need to translate and test assessment instruments in the numerous languages spoken by families. Accordingly, researchers are increasingly focused on teacher ratings to understand why educators—even those observing at the same time in the same setting—rate student behavior differently. The literature has identified *contextual*, *informant*, and *child* factors that influence teacher ratings of student behavior.

**Contextual factors.** The school context in which a teacher works may be associated with teacher ratings of student behavior. One study finds teachers working in schools with lower student-teacher ratios are more likely to rate students as more socially competent than educators working in environments with higher student-teacher ratios. In the same study, teachers working in schools with a shorter school day report fewer behavior problems and less conflict in teacher-student relationships than colleagues working in schools with a longer school day (Mashburn et al., 2006). In a separate investigation, the organizational climate of the school was associated with teacher ratings of student behavior. Teachers working in environments with positive organizational health (see Hoy & Fedman, 1987) are more likely to rate students higher on prosocial skills than educators working in schools with lower organizational health index ratings (Pas & Bradshaw, 2014). The correspondence between context and student ratings may result from the impact a teacher’s working environment has on her psychological or personal adjustment which in turn may influence her interactions with students and her perspectives on student behavior (Berg-Nielsen et al., 2011; Kraemer et al., 2003; Mashburn et al., 2006; McConaughy, 1993). More positive interactions among adults in the schoolhouse may also provide models of prosocial behavior that students adopt and display in their interactions with

both teachers and peers. Yet, the associations between contextual climate and teacher ratings are not fully understood.

**Informant factors.** Several studies find teacher psychological and personal characteristics predict student ratings. Teachers who are stressed (Berg-Nielsen et al., 2011), depressed (Berg-Nielsen et al., 2011) or burnt-out (Kraemer et al., 2003; Mashburn et al., 2006), rate students less favorably than informants with more positive psychological adjustment. The influence of the an informant's emotional state on child ratings is also observed in parent reports (Chi & Hinshaw, 2002; Gross et al., 2004; Stone, Speltz, Collett, & Werler, 2013). Mothers endorsing depressive symptoms note more problem behaviors in their children than non-depressive mothers (Barry, Dunlap, Cotten, Lochman, & Wells, 2005; Chi & Hinshaw, 2002). The quantitative investigations establish the association between an informant's emotional state and ratings, but are unable to explain if the correspondence originates with the informant's negative affect, or the child's behavior. One can imagine multiple scenarios. In one situation, a stressed, or depressed individual with high levels of unpleasant emotion views a child's behavior through a negative lens which predisposes the informant to notice problem behavior, contributing to a biased rating not truly representative of student behavior. An equally plausible scenario posits that the correspondence originates with the child's challenging behavior which fatigues, frustrates, and stresses the respondent. In the latter scenario, the informant's ratings may be unbiased and accurately represent the child's behaviors. A third scenario in which a bidirectional relationship between the child's behavior and the informant's emotional state is also tenable (see '*reciprocal determinism*' in Bandura, 1978 ). The informants' perception of the child influences the child's behavior and the child's behavior contributes to the informant's perception. Teachers, for example, behave differently toward children who they rate as having

more problem behaviors. They are more likely to reprimand children with low ratings, display negative emotions in their presence, and avoid interacting with them in social situations (Dobbs & Arnold, 2009).

A teacher's belief about her ability to manage her classroom is also associated with ratings of student behavior (Mashburn et al., 2006). Teachers who see themselves as competent in managing the classroom environment report fewer behavior problems among students than teachers who perceive themselves as less skilled classroom managers. Similarly, teachers who characterize their relationships with students as positive report fewer behavior problems than educators who report conflict in their teacher-student relationships (Berg-Nielsen et al., 2011).

Research on the influences of teachers' years of teaching experience and race on student ratings are inconclusive. Discrepant findings are reported across investigations. In a study involving 210 prekindergarten teachers, for example, investigators find that teachers with more years of experience rate students as having more problem behaviors and less social competence than their less experienced colleagues (Mashburn et al., 2006). In contrast, other publications conclude that experienced teachers rate student behavior more favorably than less experienced colleagues (Denham et al., 2009). Few studies have investigated teacher race as a factor influencing student behavior ratings, and those that have been published, report conflicting findings. Some studies find White teachers rate students more favorably than non-White colleagues (Pas & Bradshaw, 2014), others report that African-American teachers rate students more favorably than teachers from other racial/ ethnic backgrounds (Mashburn et al., 2006; Pigott & Cowen, 2000). Downey and Pribesh (2004) found that African-American teachers rated African-American students more favorably than White teachers. Additional research is needed to better understand how an educator's racial-ethnic background influences judgments about student behavior. We

need more information to understand how teachers' view students from racial or ethnic backgrounds like their own, and whether there are differences in the ways in which teachers appraise the behaviors of children from different racial and ethnic traditions.

**Child factors.** Among young children, a child's age, gender, verbal fluency, and cognitive abilities are associated with teacher ratings of student behavior. Older children and those with high scores on verbal and cognitive measures are rated more favorably than their younger classmates who may be less mature and lack the verbal reasoning skills that support social and emotional problem solving (Mashburn et al., 2006). Gender is also associated with teacher ratings of student behavior measures. Females are likely to be rated more favorably than their male peers (Alexander, Entwisle, & Dauber, 1993; Graves, Blake, & Kim, 2012; Gross et al., 2004; Mashburn et al., 2006). Additionally, boys are consistently identified as having more externalizing problems than their female age-mates (Epstein et al., 1998; Hinshaw et al., 1992; Powell, & Kamphaus, 1994; Wolcott & Williford, 2015).

The influence of a child's race on teachers' ratings may vary across developmental levels. At the preschool level, race is inconsistently associated with teacher ratings of student behavior (Graves et al., 2012; Mashburn et al., 2006). Yet, amongst children in elementary school beyond, a considerable body of research documents that teachers rate African-American students as having more problem behaviors than their White classmates (Zimmerman, Khoury, Vega, Gil, & Warheit, 1995) and African-American children are overrepresented in disciplinary referrals (Rocque, 2010). Piggot and Cowen (2000) find that both African-American and White teachers rate African-American elementary-school students as having more school adjustment problems (e.g. disruptive in class, shy or timid) and fewer competencies (e.g. peer social skills) than their White peers, even after controlling for socio-economic status.

In contradistinction to the Piggott and Cowan study, several investigations highlight the racial-ethnic congruence between rater and student as a predictive of student ratings (Downey & Pribesh, 2004; Ehrenberg, Goldhaber, & Brewer, 1995; Zimmerman et al., 1995). In an analysis of National Educational Longitudinal Study (NELS) data, researchers found that adolescents were rated more favorably on teachers' subjective measures of academic achievement when the teacher shared the same race-ethnicity as the student (Ehrenberg et al., 1995). The significant association between a rater's race/ethnicity and that of the student was replicated in a subsequent study evaluating the impact of racial-ethnic congruence on teacher ratings of student behavior. The results suggest that African-American teachers rate African-American students more favorably than White students, and White teachers rate White students more favorably than African-American students (Downey & Pribesh, 2004).

Discrepant findings are reported regarding associations between a child's socio-economic status and teacher ratings of student behavior. Some studies find no association between a child's socioeconomic status and teacher ratings of behavior (Mashburn et al., 2006); whereas, other investigations report that teachers identify more problem behaviors in students coming low SES home environments than those with higher socioeconomic status (Barry et al., 2005; Thomas-Presswood, 2008). Teachers are more likely to report disruptive behavior in males from low SES homes than males from more affluent families (Barry et al., 2005). A robust literature documents the significant impact that poverty has on families (Jensen, 2009; Thomas-Presswood, 2008). Acute chronic stress and financial instability place children at risk for social emotional challenges, cognitive delays, and health and safety issues (Jensen, 2009; McLeod & Nonnemaker, 2000). Hence, for some children, the chronic stress associated with poverty may contribute to real differences in behavior which are accurately reflected in teacher ratings.

**Instrument-related factors.** In addition to the contextual, informant, and child factors warranting further investigation, Duckworth and Yeager (2015) posit that variations in teachers' thought processes when completing rating scale measures may also influence teacher ratings. For example, teachers may interpret the behavioral indicators contained in a rating scale prompt in a multitude of ways, resulting in teachers evaluating students on distinct interpretations of a behavior. Teachers may also hold different standards and levels of stringency for specific behaviors (Edebrock, 1984; Renk & Phares, 2004). Accordingly, one student's behavior may be rated differently by teachers holding different notions of the behavior indicator contained in the prompt. Similarly, teachers may interpret the frequency responses (e.g. never, rarely, occasionally, frequently, very frequently) differently, or have a tendency to select responses in the middle of a response range rather than extreme responses resulting in inconsistency between raters (Pas & Bradshaw, 2014).

Although the possibility for rater effects derived from differences in cognitive processes when responding to standardized survey items is well documented in the research literature on survey design, the prevalence and nature of these effects in teachers' ratings of students' social and emotional competencies is unknown. It is unclear to what extent teachers' ratings of students' competencies are influenced by differences and inconsistencies in teachers' thought processes in response to the measure itself. Teachers may disagree on how particular behavioral terms are defined, the contexts in which the behavior--as understood by the rater--can be observed, and the ways in which observations of a single student contribute to the selection of response. For example, two respondents comprehending an item about the frequency of paying attention, may hold different notions about what it means to pay attention, and may be basing their assessments of 'paying attention' on different indicators of behavior. These differences,

originating from variations in the ways in which respondents interpret the language on the measure, may result in teachers rating the same behavior according different internally held interpretations and standards for behavior. Differences associated with teachers' comprehension, retrieval, and responses to the measure itself will be referred to as *instrument-related rater factors*.

### **Research Questions**

Past studies investigating the associations between teacher ratings and factors independent of child behavior suggest possible explanations for differences in teachers' ratings of students' behavior and social emotional competencies. Additional research is needed to explore how these and other 'rater effects' influence teachers' ratings of student behavior and social-emotional competencies. Data gleaned from these measures are being utilized to evaluate the efficacy of SEL interventions (Durlak, Weissberg, & Pachan, 2010; Greenberg et al., 1995), school quality ("The Every Student Succeeds Act," 2015), and the emotional well-being of students (Duckworth & Yeager, 2015). In some school districts, teachers' evaluations of students' social and emotional skills are becoming part of students' permanent academic record (Duckworth & Yeager, 2015; Elias et al., 2016). These measures have far reaching consequences and potentially may impact future educational opportunities, employment, or determinations about suitability for roles and responsibilities. Urgent attention is needed to understand how teachers experience, interpret, and think about student behavior in the evaluation of social and emotional competencies to learn how teachers' thoughts influence rating scale responses.

We know little of how teachers' thoughts impact their evaluations of students' social and emotional skills. To date, factors associated with teachers' ratings have been studied solely from

a quantitative orientation, and these investigations have focused on concurrent factors present at the time of rating. Prior research has not investigated how teachers' experiences, thoughts, and beliefs influence their ratings of student behavior. Published studies identify the existence of relationships between various factors and teachers' ratings; yet work remains to explain *how* these factors influence teacher ratings.

The primary goal of this project is to understand *how* teachers form judgments about students' social and emotional competencies. The study sets out to investigate the ways in which child, informant, contextual, and instrument-related factors influence teachers' ratings of students. The following research questions guided the study design and analysis:

- (a) What is the relationship of within classroom variability to between-classroom variability in teacher ratings of students' social and emotional competencies?
- (b) To what extent are child, informant, contextual, and instrument-related factors associated with teachers' ratings of students' social and emotional competence?
- (c) How do child, informant, contextual, and instrument-related factors play into teachers' thinking when rating students' social and emotional competencies?
- (d) How do teachers expressed thoughts when completing a social and emotional rating scale reflect child, informant, contextual, and instrument-related factors in the process of rating students' social and emotional skills and competencies? How do these factors influence a teachers' ratings of individual children?
- (e) How do teachers' interpretations and responses to a social and emotional rating scale measures compare among teachers working in the same and different school contexts?

### **Chapter 3: Conceptual Model and Theoretical Framing**

Teacher rating scales (e.g. Behavior and Emotion Rating Scale, 2<sup>nd</sup> Edition, Social Skills Improvement System–Rating Scale, Devereux Student Strengths Assessment), like the DESSA-Mini, include a series of items that prompt teachers to rate how frequently a student exhibits a behavior. The seemingly simple act of reading behavior descriptions and assigning ratings for how frequently a student displays certain behaviors is--in actuality--a multi-faceted activity that places numerous cognitive demands upon the respondent. To rate the behaviors of each student, the teacher must read and understand the literal language contained in the measure. She must also understand the significance of words used in the directions, the prompt, and the frequency labels. The informant must also connect this literal language to her own practical understanding of what is meant by each term. The respondent is subsequently called upon to recollect personal knowledge of students and experiences with each child over a specified period.

Each of the discrete cognitive tasks typically completed when rating students' competencies may be influenced by factors independent of student behavior. As previously noted, prior research finds that--in some instances--child, informant and contextual factors are associated with teacher ratings. Yet, the acceptance of teachers' ratings as measures of 'true differences' in student behavior hinges on the implicit assumption that teachers objectively and consistently complete a shared set of cognitive tasks, and in doing so, draw upon shared behavioral standards to complete ratings. Interpreting teacher ratings as accurate measures of students' social and emotional competencies presumes that teachers are universally adept, unbiased, and accurate in comprehending rating scale items, retrieving relevant memories, and assigning ratings in a similar manner. In the absence of evidence supporting these assumptions,

individuals interpreting rating scale results risk drawing erroneous conclusions about students and programs that may have far reaching consequences (Duckworth & Yeager, 2015).

To better understand teachers' ratings of students' social and emotional skills and competencies, the research community must investigate how teachers interpret the language on standardized measures and the ways in which teachers' personal experiences, thoughts, and beliefs--across time and contexts--influence their thought processes and perceptions of behavior when rating students. This mixed methods study is based on a sequential explanatory design (Creswell & Plano Clark, 2011) in which a qualitative investigation follows the collection of quantitative teacher ratings. In the quantitative component, teachers completed ratings of social and emotional competency for all elementary students within the participating district. The qualitative investigation attempts explain these ratings by exploring teachers' internal thought processes. The project relies on concepts from the bio-ecological, social learning, and cognitive aspects of survey methodology theoretical traditions to frame the qualitative investigation.

### **Theoretical Bases for Development of Thoughts, Beliefs, and Expectations about Behavior**

The development of teachers' thoughts, values, and beliefs about student behavior may be usefully studied through bio-ecological (Bronfenbrenner, 1995; Bronfenbrenner & Morris, 1998) and social learning theories (Bandura, 1977b). In accordance with bio-ecological theory, a teacher's individual heritable characteristics, such as race, gender, and temperament, interact with the multi-level contexts that she experiences. Her interactions within these multiple contexts shape her thoughts, values, beliefs over time (Bronfenbrenner, 2005; Bronfenbrenner & Ceci, 1994).

The bio-ecological model highlights three interdependent characteristics of experience that guide my understanding of how teachers' perceptions of student behavior develop. These

characteristics include: (a) *person characteristics*, (b) *time characteristics*, and (c) *context characteristics*.

**Person characteristics.** Bio-ecological theory attends to the influence an individual's heritable characteristics (e.g. gender, race/ethnicity, capabilities, temperament) have on the developmental processes. Per the model, bio-psychological characteristics influence the experiences that individuals have in the world. A person's gender, for example, may influence the behavioral expectations that others hold for an individual. In some contexts, males may be praised for speaking out, using body language to assert themselves, and speaking loudly. Yet, these very same behaviors may be discouraged in females, even within shared contexts. Individuals of different genders may be also be differentially reinforced or punished for the exhibiting the same behaviors. The differential responses may be partially attributable to the interactions of individual person characteristics—in this case, gender--and the social norms and values held by those in a context. One can imagine similar response differences directed towards individuals based on race, or temperament. The bio-ecological model posits that heritable person characteristics influence the interactions that individuals have in their daily experiences, and these interactions shape personal development over time (Bronfenbrenner, 2005; Bronfenbrenner & Morris, 1998).

As such, a teacher's experiences—shaped by person characteristics of gender, race, temperament, and abilities—influence her personal development and values held for specific behaviors. Additionally, a teacher's person characteristics also may inform the ways in which she experiences and perceives student behavior. To consider how person characteristics may inform teachers' perceptions and judgment of student behavior, imagine two educators with different temperament qualities teaching in the same classroom. One teacher is introverted,

prone to negative affect, and has poor emotion regulation skills. The other is extroverted, experiences low negative affect, and has high emotion regulation abilities. These two individuals may respond to the same student behavior differently and may make different judgments about student skills because of their unique person characteristics.

**Time characteristics.** The bio-ecological model conceptualizes time as having a multi-faceted influence on development and happening on multiple levels. *Micro-level* time includes the immediate interactions that individuals experience at a point in time. A teacher resolving a conflict between students in her classroom experiences that exchange at the micro-level. Her interaction is happening in present time with students in her immediate environment. Although daily exchanges occur at the *micro-time* level, these encounters are informed by the interactions that an individual has experienced over an extended period. *Macro-time* is a term that is used to describes changes that happen over broad spans of time, such as changes in societal norms, which may result in significant changes to norms governing social behavior.

**Context characteristics.** Perhaps the most defining feature of the bio-ecological model and the one that I will draw upon most heavily for this study is the conceptualization of the multi-level contexts that shape individual development. The model describes an individual's context as a set of nested ecological systems in which the individual interacts with the external world. The impact of these interactions is bi-directional and the individual and the environment shape one another.

Bio-ecological theory purports that an individual's context is comprised of four interdependent ecological levels: (a) *micro-system*, (b) *meso-system*, (c) *exo-system*, (d) *macro-system*. The *micro-system* includes the local contexts in which the individual interacts daily. For each teacher in the current study, for example, her current micro-system would include the

classroom in which she teaches, her home, and other places she visits on a regular basis (e.g. church). The *meso-system* represents the space in which interactions occur between the multiple elements in the micro-system. The *exo-system* is more distant from the individual's daily experiences and includes levels of influence that do not have a direct impact on the individual. Changes to district educational policies, or state educational leadership, for example, may be thought to happen at the *exo-system* level. Although events occurring at the *exo-system* level may not have a direct and immediate impact on a teacher's daily school experiences, changes in the *exo-system*, may contribute to future microsystem changes that impact testing requirements, or class size, thereby altering a teacher's daily experience. The *macro-system* level includes broad reaching influences such as cultural ideologies, federal education policies (e.g. NCLB, Common Core, racial integration of schools), or pervasive social beliefs (e.g. value for cultural diversity, transgender rights). These factors at the macro-system level may influence and shape experiences that transpire at all levels of the nested system and often contribute to changes that impact individuals over generations.

In explaining how person characteristics, context characteristics, and time characteristics influence individual development, the bio-ecological theory asserts that individual behavior must be understood across time and contexts. An individual is shaped by her heritable characteristics and her daily experiences environment; however, she is also influenced by distant experiences that transpire in different locations at different points in time. In the current study, I rely upon the bio-ecological theory of development as a framework for understanding how person, time, and context characteristics influence individual experiences, and how these characteristics contribute to the knowledge, beliefs, and expectations held for behavior.

## **Theoretical Bases for the Formation of Teacher Expectancies about Student Behavior**

Social learning theory helps lay the theoretical groundwork teachers' notions of student behavior. The theory emphasizes the importance of modeling as a primary driver of individual behavior and postulates that individuals adopt or reject the behaviors observed in their environment, partially in regard for their judgment of the model and the responses elicited by particular behaviors (Baer & Bandura, 1963; Grusec, 1992; Sears, 1951). Rooted in the field of information processing, social learning theorists assert that individuals engage in internal thought processes to critically evaluate the behaviors they observe in the external world. Individuals selectively attend to the behaviors of others, interpret and make meaning from their observations, and develop internalized working models of behaviors that attach meaning and value to the behavior. This process sometimes requires individuals to reconcile diverse behaviors modeled by a single individual (e.g. inconsistent responses to a demonstrated behavior), or across individuals modeling different behaviors in similar contexts. Observations can occur either directly or indirectly through visual or print media (Bandura, 1969). The resulting internal models of behavior are then stored in long term memory and referenced to make sense of new experiences (Bandura, 1969, 1977b; Grusec, 1992).

The value individuals ascribe to behaviors --like the behaviors themselves—are also socially learned through observing the evaluative judgments made by others in the environment. Behaviors that are punished or rejected by people who are respected, or who control rewarding resources, are more likely to be shunned by the observer (Baer & Bandura, 1963). In contrast, behaviors that receive praise and are reinforced by respected others are more likely to be valued and imitated.

Research has shown that individuals moral judgments about behaviors that are ‘right’ or ‘wrong’ are similarly learned. In experiments conducted by Bandura and McDonald (1963), children were asked to share their judgments about particular behaviors. They were then exposed to adult models who expressed moral judgments that aligned with their own, or whose judgments contradicted those expressed by the child. Following exposure to the adult responses to the referenced behaviors, children were then presented with another opportunity to evaluate the same behaviors. Researchers found that children substantially altered their moral judgment to align with judgments modeled by adults.

The social learning theory, grounded in the broader bio-ecological developmental framework, suggests the need to consider how the behavioral models that teachers observed during their personal development—in school, at home, and in their communities—influence their current judgments about student behavior. These theoretical traditions speak to the influences of the intersecting relationships among the teacher, multiple ecosystems, and others in shaping her understanding of the world. It is these personal experiences and internalized mental representations of behavior that teachers draw upon when assessing their students’ behavior. These models held in long term memory, are activated when a teacher is prompted to respond to questions about student behavior.

### **Theoretical Bases for Teachers’ Cognitive Responses to Rating Scale Measures**

The content that populates teacher judgments of student behavior is activated and evaluated through a series of cognitive tasks. I draw on the literatures in cognitive aspects of survey methodology (CASM) and information processing to conceptualize teachers’ thought processes when rating students’ social and emotional competencies. CASM is rooted in information processing theory and the conceptualization of human thought as driven by internal

processes, analogous to those within computer systems (Reitman, 1964). At the simplest level, the information processing model posits that individuals selectively attend to external sensory information, and store this information in *sensory memory*, or the sensory register. Information is then either lost, or transferred to *working memory*, sometimes called short-term memory. Information received in working memory is compared with information in long-term memory to make meaning of the new experience. The newly encoded information is then either lost (through ‘decay’) or stored in *long-term memory* (Atkinson & Shiffrin, 1968). Thus, according to information theorists, human thought and behavior can be explained through a careful and detailed analysis of cognitive processing tasks (Newell & Simon, 1961). If one can correctly identify and define the internal thought processes enacted following the input of sensory experiences, one can explain human behavior.

Building on the information processing conception of human thought processes, CASM researchers mapped the cognitive tasks that respondents complete when responding to surveys. In this study, I utilize the CASM cognitive response model (Tourangeau, 2000; Tourangeau, R., 1984; Tourangeau & Rasinski, 1988) as a conceptual tool for investigating teachers’ thought processes when completing rating scale measures. Although the cognitive response process model was developed to describe cognitive processes enacted when completing surveys, teachers engage in a comparable set of tasks when rating students’ social and emotional competencies. Teachers read items and respond to prompts by selecting a rating from the available response options. The cognitive response model provides a useful frame for analyzing the ways in which educators think about and respond to rating scale measures.

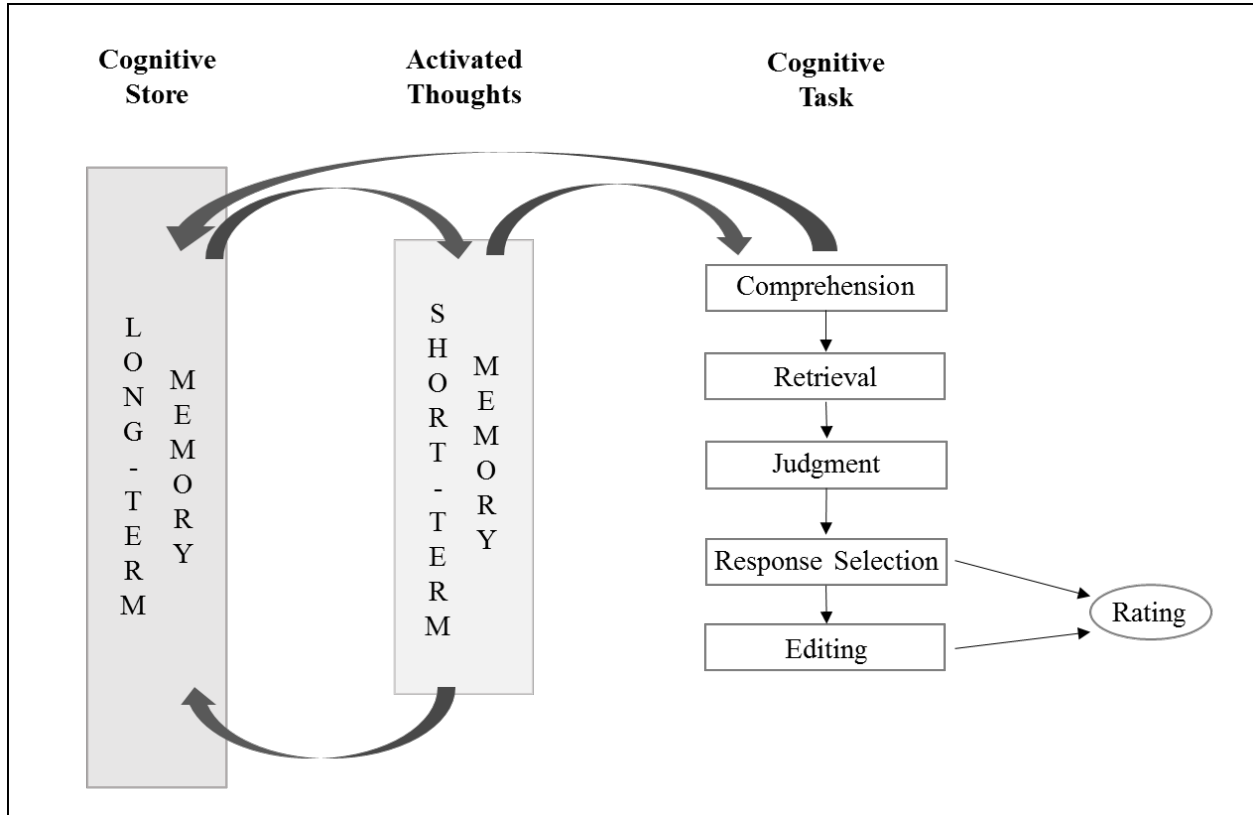
Tourangeau et al. identified 4 cognitive tasks that individuals complete when responding to survey questions (See Table 3.1).

**Table 3.1.** *Components of the Response Process*

Component	Specific Process
Comprehension	Attend to questions and instructions Represent logical form of question Identify question focus Link key terms to relative concepts
Retrieval	Generate retrieval strategy and cues Retrieve specific, generic memories Fill in missing details
Judgment	Assess completeness and relevance of memories Draw inferences based on accessibility Integrate material retrieved Make estimate based on partial retrieval
Response	Map judgment onto response category Edit response

Reproduced from Psychology of Survey Response, Tourangeau, Rips & Rasinski (1988, p.8)

These four tasks include the *comprehension* of the prompt, *retrieval* of pertinent information from long term memory, *judgment* about retrieved information, and the *selection of a response* option from the choices provided. According to the model, respondents complete these tasks in sequential order; however, the combination and ordering of these processes may vary based on a variety of factors including the respondent's motivation to produce an accurate and quick response (Tourangeau, 2000; Tourangeau & Rasinski, 1988). The four processes are described in more detail below and illustrated in Figure 3.1.



**Figure 3.1.** *Cognitive response to rating scale measure*

**Comprehension.** Teachers asked to respond to a rating scale must first *comprehend* the behavior referenced in a question or prompt. In doing so, the respondent reads words in the question, assigns meaning to the specific terms by eliciting personal definitions of the terms, and infers what is meant by the term in the context of the question. This process involves the activation of information about the specific terms from long term memory and possibly subconscious personal notions about behaviors, ideas, or experiences held by individual informants.

Research has shown differences in individuals' experiences over time contribute to differences in the meaning attributed to specific words which, in turn, may lead to multiple interpretations of items that appear on a rating scale. CASM researchers have investigated how

the language in surveys (e.g. context, syntax, word choice) and the ordering of items impact comprehension (Norman M. Bradburn, 2004; Sudman, 1982), but have only more recently paid attention to differences in the ways in which individuals interpret seemingly unambiguous terms (Schwarz, 2007). Desimone and Floch (2004), for example, investigated how academic researchers, teachers, and principals interpreted survey terminology pertaining to national school reform. These researchers discovered that teachers interpreted reform terminology (e.g. rewards and sanctions, professional development, high-stakes assessment) in the context of their daily teaching practices, rather than in the broader reform context in which the survey developers intended (Desimone & Le Floch, 2004; Wilhelm & Andrews-Larson, 2016). Comprehension differences, stemming from respondents' personal experiences in specific roles and contexts, resulted in teachers essentially answering different questions than were intended by survey developers.

**Retrieval.** Following the reading of items, teachers activate information relevant to the behavioral indicators identified in the prompt, and *retrieve* examples of behaviors. While holding the mental representation of the queried behavior in working memory (e.g. paying attention), the respondent retrieves memories of the student being rated. The recall, or retrieval, task can be thought of as a sampling activity in which the, recency (Higgins & Kruglanski, 1996; Higgins, Rholes, & Jones, 1977), frequency (Tourangeau, 1984), emotional intensity (Brown & Kulik, 1977; Holmes, 1970), and emotional valence (Holmes, 1970; Jersild, 1931; Meltzer, 1931) of an event may influence access to a memory and the likelihood of that event being sampled during the retrieval process. Emotionally charged events, experiences that contradict expectations (Bower, Black, & Turner, 1979), and recent events (Broadbent & Broadbent, 1981; Murdock,

1963, 1999) may be particularly salient to respondents and more easily retrieved (Tourangeau, R., 1984).

**Judgment.** It is important to note that the retrieval process does not necessarily result in a *judgment* as the respondent must evaluate whether the retrieved information is accurate, and assess whether the information accessed is representative of the respondent's overall experience of a student. Difficulty in retrieval, and gaps in memory, may contribute to the use of additional retrieval strategies intended to supplement recollections with additional relevant information stored in long term memory. In some instances, respondents rely on 'top down' assessments of an individual, or experience (Ontario Symposium on Personality and Social Psychology, 1981; Tourangeau, 2000) and attribute behaviors to an individual based on a global impression of the student.

Researchers have reliably replicated 'halo effects' (Thorndike, 1920) in which evaluators judge skills, or attributes, based on non-relevant generalized biases even in the presence of relevant information about the skill or attribute of interest (Babad, Inbar, & Rosenthal, 1982; Nisbett & Wilson, 1977). Negative halo affects have also been observed in teacher ratings of students, particularly among students who display oppositional or aggressive behaviors (Abikoff, Courtney, Jr, & Koplewicz, 1993; Blunden, Spring, & Greenberg, 1974; Siegel, Dragovich, & Marholin, 1976). Teachers negative perceptions of one behavior (e.g. aggression, defiance) can spill over into ratings of other behaviors (e.g. hyperactivity, impulsivity, attention, etc.), painting a more negative representation of a student than may be accurate, resulting in a more extreme unfavorable rating (Abikoff & Gittelman, 1985).

In some instances, respondents recall memories that provide conflicting information about behaviors, attitudes, or experiences. Many theories have been advanced to explain how

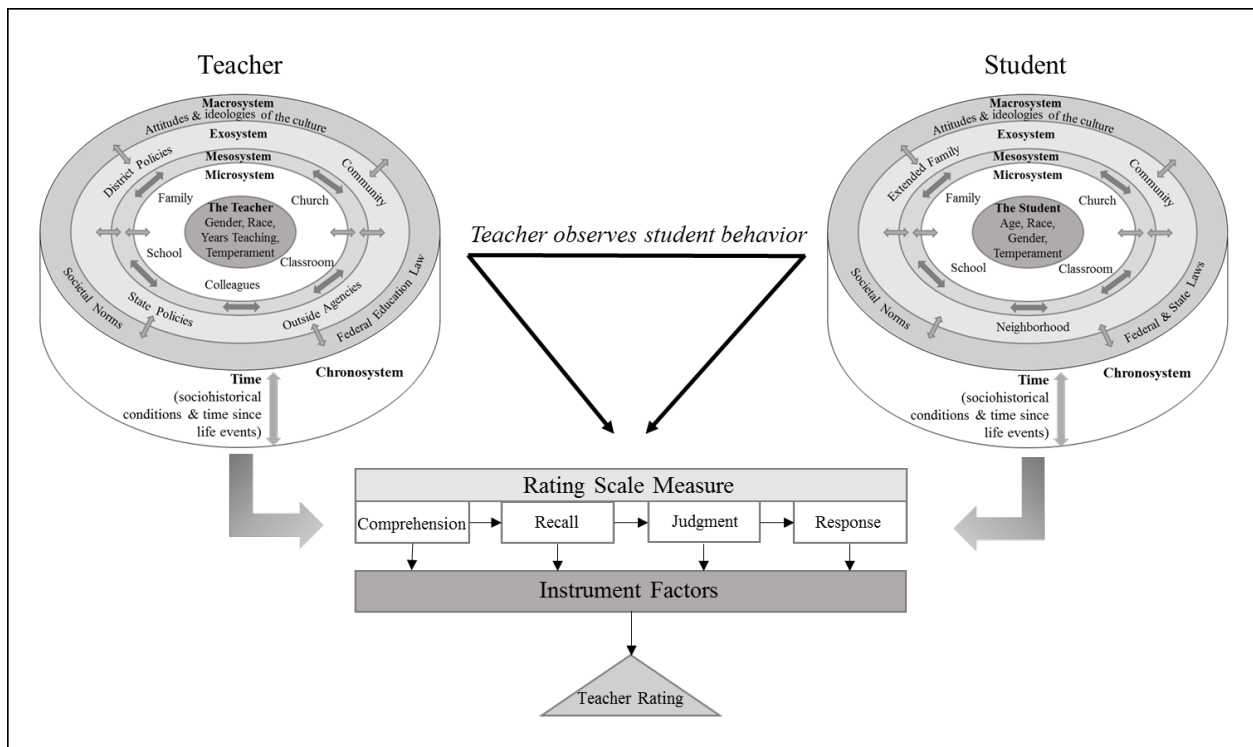
respondents reconcile inconsistent recalled information. Some respondents elect to focus on a single memory to support a judgment, others average, or sum, experiences over some period to form a judgment (Anderson, 1981; Tourangeau, 2000). Numerous heuristics (e.g. Tversky & Kahneman, 1974) have been advanced in the CASM literature to explain the ways in which individuals arrive at judgments (see Hastie, 1983 for a comprehensive review).

**Response Selection.** The selection of a response from the available response choices can be thought of as a mapping of judgment onto the response. When considering the possible response options, the respondent must comprehend the language used in the response options. In the case of the present study, teachers must interpret the frequency ratings employed by the measure. What does it mean to ‘frequently’ pay attention? How does a child who ‘frequently’ pays attention look differently from the child who ‘occasionally’ pays attention?

Once the teacher has selected the response that best matches her judgment, the selection may be edited by the respondent in order to achieve moderation, consistency, or social desirability (Tourangeau & Rasinski, 1988). Some individuals may feel the need to strive for consistency across responses, or to balance previous responses to achieve moderation. Teachers may strive to have a consistently positive appraisal of an individual student. If the rater responds with either an extreme positive, or negative response, for an item, she may try to balance an extreme response with an equally extreme response in the opposite direction, or with more moderate responses for the remaining items. Social desirability may also affect teacher ratings. Teachers who want to appear kind or tolerant, may rate a student more favorably than she would if she were not attending to how she is perceived by those who access to her ratings.

The cognitive response model defines, sequences, and describes a proposed set of cognitive tasks that individuals complete to form a judgment in response to a question. These

cognitive processes of *comprehension, retrieval, judgment, and response selection* draw upon the memories, knowledge, beliefs, and biases held by individual teachers and developed because of personal experiences with others over time and across multiple contexts. In this investigation, I utilize concepts from the CASM, social learning, and bio-ecological theoretical traditions to frame the qualitative investigation of factors contributing to teachers' ratings of students' social and emotional competencies (See Figure 3.2.).



**Figure 3.2.** *Factors influencing teacher ratings*

I rely on the bio-ecological tradition to explain how an individual's experiences over time and contexts contribute to the development of individuals' thoughts, beliefs, and biases. More specifically, I highlight the importance of social modeling as a primary mechanism influencing teachers' expectations and perceptions of student behavior. A teacher's experiences, learned values, and standards for student behavior are activated as she engages in the cognitive tasks of responding to a rating scale measure. The proposed framework hypothesizes a set of cognitive

tasks and the cognitions that are activated as teachers observe and judge student behavior. The framework is constructed to guide the investigation into how teachers' thoughts during assessment contribute to rater effects that impact teacher ratings of student behavior.

## **Chapter 4: Method**

The present study emerged from work that began in the fall of 2013, when I was employed as the Social and Emotional Learning Curriculum Developer for an urban school district in the Pacific Northwest. In this role, I partnered with the Director of Research and Accountability and faculty from the University of Washington to develop and implement an assessment plan to measure students' social and emotional competencies to evaluate the impact of universal SEL curricula on students' skills. The 8-item DESSA-Mini teacher rating scale was selected for its efficiency, alignment with the CASEL competencies, and strength-based orientation. The DESSA-Mini parenting rating scale was also considered; however, its availability in only two languages (e.g. English, Spanish) would have severely restricted parent participation, as more than 80 languages are spoken by families enrolled in district schools.

In 2014-2015, the DESSA-Mini was piloted at three elementary schools purposefully sampled to collectively represent the racial-ethnic and socio-economic diversity of students enrolled in the district. All classroom teachers in each of the selected schools rated each of their students during the fall and spring terms. At each time point, teacher ratings yielded significant inter-classroom mean score differences, even among informants teaching the same grade level in the same school. Mean score differences were observed between male and female students, and among racial and ethnic groups. The rating scale results raised important questions about whether teachers' ratings reflected true differences in students' social and emotional classroom behaviors and competencies, or if the score differences could be explained by other factors.

In the spring of 2015, the district required all elementary classroom teachers in each of the 17 elementary schools to rate students' social and emotional skills using the DESSA-Mini Rating Scale. Classroom mean score differences among same grade classrooms within the same

school continued to be observed. The findings from teachers' ratings continued to generate numerous questions about the differences in teachers' ratings and the differences observed between groups of students based on racial/ethnic background, free/ reduced lunch status, and eligibility for English language learning services.

### **Study design**

To investigate how teachers' thought processes contribute to teacher ratings of students' social and emotional competencies, I utilized a sequential mixed methods explanatory design (Castro, Kellison, Boyd, & Kopak, 2010). The quantitative analysis of student scores is followed by a qualitative investigation designed to help explain quantitative results. The combination of qualitative and quantitative methodologies leverages the complementary strengths of each research tradition. The quantitative strand of the study generated numerical measures of student social and emotional competency at the student, classroom, and school level.

The qualitative component of the project investigated the quantitative findings using cognitive interviews to explore teachers' thought processes. The semi-structured interviews were designed to elicit teacher's thoughts to help explain *how* child, classroom, school, and instrument-related factors influence teacher ratings. In leveraging the complementary strengths of quantitative and qualitative research methods, the project provides a more holistic understanding of teachers' ratings.

**Setting.** The study was conducted in a mid-size (approximately 20,000 students) urban minority-majority school district in the Pacific Northwest. More than 80 languages are spoken by children and families enrolled in the district, and 34% of children speak a first language other than English. Approximately 19% of the students were eligible for free or reduced-price school meals. Students within the district are racially and ethnically diverse. Approximately 3% of

students identified as Black/African-American, 36% of students were Asian, 12% Hispanic/Latino, 9% Multi-Ethnic, and 40% White. English Language Learning services were provided to 12% of students. The average elementary class size was 22.9 students.

The district employs approximately 1,100 classroom teachers. More than 380 teachers held National Board-Certification and over 75% had earned a master's degree. Elementary classroom teachers in each of the 15 non-language based elementary schools taught a common curriculum. The two language immersion schools provided a modified curriculum to accommodate language instruction.

**Participants.** All grade 3-5 certificated non-special education classroom teachers (N = 173) rated their students as part of a district initiative. Ratings were gathered from 61 grade 3 classroom teachers, 59 grade 4 classroom teachers, and 53 grade 5 classroom teachers working in 17 elementary schools across the district. In total, classroom teachers completed 4,005 ratings for students in grades 3-5. The gender and racial and ethnic backgrounds of students rated by teachers was consistent with district demographics: Female (48.8 %), White (38.7%), Asian (37.7%), Hispanic/ Latino (12.1%), Multi-Ethnic (8.3%), Black/African-American (2.5%), American Indian/ Alaska Native (<1%), Hawaiian/ Pacific Islander (<1%).

### **Quantitative Component Methods**

**Outcome Measure.** Students' social and emotional competency was assessed using the 8-item DESSA-Mini (Naglieri et al., 2009). The DESSA-Mini is a strength-based rating scale comprised of items selected from the 72-item Devereux Student Strengths Assessment (DESSA) (LeBuffe & Naglieri, 2008). The DESSA is grounded in a resiliency framework and was designed to screen and track within-child behaviors associated positive cognitive, emotional, and social development (Nickerson & Fishman, 2009). The 8 conceptually derived scales that

comprise the DESSA are aligned with the 5 CASEL competencies and include 3 additional competency areas (e.g. Personal Responsibility, Optimistic Thinking, Goal Directed Behavior). The DESSA subscales include: Self-Awareness (understanding of strengths and limitations), Social Awareness (interacting respectfully with others), Self-Management (controlling emotions and behavior), Goal Directed Behavior (persisting and completing tasks), Relationship Skills (maintaining positive connections with others), Personal Responsibility (being careful and reliable in actions, contributing to group efforts), Decision Making (using personal values to guide choices and actions), and Optimistic Thinking (being hopeful and positive in thinking). The DESSA subscale and composite scores are convergent with related competency scales on other psychometrically tested instruments (e.g. Behavior Assessment System for Children, Second Edition—BASC 2, Behavioral Emotional Rating Scale, Second Edition—BERS-2) and are divergent from scales tapping maladaptive social and emotional behaviors (Nickerson & Fishman, 2009). Sample DESSA items contributing to each subscale are shown in Table 4.1.

**Table 4.1.** *DESSA Scales and Sample Items*

Scale	Sample items
Self-Awareness (7 items)	Gives an opinion when asked Describes how he/she is feeling
Social Awareness (9 items)	Gets along with different types of people Cooperates with peers or sibling
Self-Management (11 items)	Waits his/ her turn Stays calm when faced with a challenge
Goal Directed Behavior (10 items)	Keeps trying when unsuccessful Takes steps to achieve goal
Relationship Skills (10 items)	Offers to help somebody Compliments or congratulates somebody
Personal Responsibility (10 items)	Remembers important information Handles belongings with care
Decision Making (8 items)	Accepts responsibility from what he/ she did Follow example of positive role model
Optimistic Thinking (7 items)	Says good things about himself/ herself Looks forward to classes or activities

For each item, teachers rate how frequently the student exhibits these behaviors on a 5-point Likert-type scale (0 = *never*, 4 = *very frequently*). Teacher ratings yield 8 scale scores and a Social Emotional Competency (SET) composite score.

The DESSA-Mini, developed after the DESSA, retained a focus on behaviors that support resilience. In addition, the DESSA-Mini intentionally targets behaviors and competencies emphasized in social and emotional learning (e.g. Payton, 2009; Zins, Bloodworth, Weissberg, & Walberg, 2007) and positive youth development frameworks (e.g. Catalano, Berglund, Ryan, Lonczak, & Hawkins, 2004). The DESSA-Mini is available in 4 equivalent

forms that are derived from unique combinations of individual DESSA items and highly predictive of the DESSA full-version Social-Emotional Composite (SEC) (Naglieri, LeBuffe, & Shapiro, 2009). Importantly, the four DESSA-Mini forms do not include a designated item for each of the 8 competencies evaluated by the DESSA full version. Instead, the 8 items on each form—in combination—are highly correlated with the DESSA social and competency (SEC) total score. Hence, individual scale scores are not generated on the DESSA-Mini measure. A single Social and Emotional Total (SET) t-score is calculated. T-scores between 28-40, inclusively, are classified in the ‘need’ range. Scores of 41-59 comprise the ‘typical’ range and scores of 60 and above make up the ‘strength’ range.

The DESSA-Mini 2005-2006 norming sample of K-8 children (N = 1250) approximated the population of U.S. K-8 students at that time with respect to gender, age, geographical region, and socio-economic status (24.2% eligible for free or reduce price lunch). The sample was comprised of American Indian/Alaska Native (n = 8, .8%), Asian (n = 24, 2.5%), Black / African-American (n = 371, 38%), Native Hawaiian/ Pacific Islander (n = 4, .4%), Hispanic (n = 303, 24.3%) and White (n = 570, 58.3%) students. Differences between groups were evaluated using *d*-ratios. Values below 0.2 represent negligible differences and values of 0.2 - 0.5 characterize small differences (Jacob Cohen, 1987). On the form used in this study (form 1), small differences were observed between Black/African-American and White students (*d-ratio* = 0.26) and between Hispanic and Non-Hispanic students (*d-ratio* = 0.24). Moderate differences between males and females were also found (*d-ratio* = 0.42). The DESSA-Mini technical manual explains gender differences as reflective of true gender disparities in social and emotional competency that exist in K-8 children (Naglieri et al., 2009). Test-retest reliability for teachers rating the same child on two separate occasions within an interval of 4-8 days ranged

from 0.88 to 0.94 ( $r = 0.88 - 0.94$ ) across the 4 forms. Inter-rater reliability between educators observing a child in the same environment ranged from 0.70 - 0.81 ( $r = .70 - .81$ ).

**Procedure.** The DESSA-Mini was a district required assessment that teachers completed as part of negotiated contract agreement. To maintain standardized administration procedures, instructional leaders coordinating the completion of teacher ratings were directed not to provide additional information on interpreting behavior indicators, or guidance on the frequency ratings. An hour of staff meeting time was allocated for teachers to rate their students. Teacher ratings were completed in October 2015.

**Data preparation.** The dependent variable is a teacher rating of students' Social and Emotional Total competency (SET) as measured by the DESSA-Mini SET t-score. Individual student SET t-scores were utilized as the sole outcome measure.

*Independent quantitative variables.* The child predictors were gender, age at rating, race/ethnicity, English language learning status, special education status, disability accommodations status, and gifted services status. Student gender was dummy coded with 0 for male and 1 for female. The age at rating was entered as a student's age in months at the time of rating (October 2015). Race and ethnicity were dummy coded for each of the federal race and ethnic heritage codes. ELL status was also dummy coded with eligible for service coded as a 1 and ineligible coded as a 0. Similarly, eligibility for special education services, disability accommodations, and gifted services were coded with 1 for eligible and 0 for ineligible. Contextual covariates were also entered as predictors to determine if classroom level characteristics were associated with teacher ratings of students' social and emotional competency. Classroom variables included class size and the percentage of students in a

classroom coming from non-White non-Asian racial/ ethnic groups (percent non-White and non-Asian students). Each of these variables was entered as a numerical value.

School level predictors included school enrollment numbers as of October 2015, the average years of teaching experience amongst the teaching staff, the mean percentage of grade 3-5 students passing the Smarter Balanced Assessment state test of English Language Arts (spring 2016), and the mean percentage of grade 3-5 students passing the Smarter Balanced Assessment state test of Math (spring 2016). The percent of students eligible for meal subsidies was also entered as school covariate. Each of the school level predictors was entered as a numerical value. Dependent variables are summarized in Table 4.2.

**Table 4.2.** *Summarization of Independent Quantitative Variables*

Variable	Summary
Gender	Coded 0 for male and 1 for female
Race/ethnicity	Federal Race/Ethnic groups dummy coded for each federal race/ ethnic group
English language learners	Student eligible for English language services. Eligible coded as 1 and ineligible coded as 0
Special education eligible	Student eligible for special education services. Eligible coded as 1 and ineligible coded as 0
Disability accommodations	Students eligible for disability accommodations who do not receive special education services. Eligible coded as 1 and ineligible coded as 0
Gifted services	Student eligible for gifted services. Eligible coded as 1 and ineligible coded as 0
Percent underserved students	Percent of non-White, non-Asian students in classroom
Class size	Number of students in classroom at time of DESSA-Mini rating
Percent passing English	Mean percent of grade 3-5 students in school passing 2015-2016 Smarter Balance English language arts exam which is aligned to Common Core learning standards
Percent passing math	Mean percent of grade 3-5 students in school passing 2015-2016 Smarter Balance Math exam which is aligned to Common Core learning standards
Average years teaching experience	Average years of teaching experience among teachers in school
Enrollment	Number of students enrolled based on October 2015 count
Percent free/reduced	Percent of students in school receiving free or reduced price meals as of May 2016

**Data analysis.** Descriptive statistics were calculated to evaluate mean SET t-scores at the district, school and classroom level. A separate descriptive analysis was conducted to evaluate ratings completed by the 15 participating teachers. Hierarchical Linear Modeling (HLM) (Raudenbush, 2002) techniques were used to evaluate sources of within and between-classroom and school variance in teachers' ratings of students' social and emotional competency in the full sample, and in Title 1 and Non-Title 1 school sub-samples. Zero-order correlations were calculated to evaluate collinearity among variables. Simple baseline HLM models, equivalent to one-way ANOVAs, were developed to examine the contributions of classroom and school differences to teachers' ratings of individual students. In equation 1a, a teacher's rating ( $Y$ ) for a student ( $i$ ) in a classroom ( $j$ ) and school ( $k$ ) is a function of the teacher's mean classroom score ( $\pi_{0jk}$ ) and measurement error ( $e_{ijk}$ ) associated with random child factors. The second level of the model (1b), indicates that each teacher's classroom mean score ( $\pi_{0jk}$ ) is a function of the grand mean across schools ( $\beta_{00k}$ ) and the errors associated with score differences between classrooms ( $r_{0jk}$ ). Equation 1c states that school mean scores are a function of the grand mean of all schools ( $\gamma_{00}$ ) and random error between-schools ( $u_{0jk}$ ). Equation 1d represents a mixed model that predicts a student's score as a function of the overall grand mean and error associated with random effects at the child ( $e_{ijk}$ ), classroom ( $r_{0jk}$ ) and school ( $u_{0jk}$ ) levels.

$$Y_{ijk} = \pi_{0jk} + e_{ijk} \quad (1a)$$

$$\pi_{0jk} = \beta_{00k} + r_{0jk} \quad (1b)$$

$$\beta_{00k} = \gamma_{00} + u_{0jk} \quad (1c)$$

$$Y_{ijk} = \gamma_{00} + r_{0jk} + u_{0jk} + e_{ijk} \quad (1d)$$

To estimate the variance attributable to student, classroom, and school factors, intra-class correlations were computed for each of the three samples. The total variance ( $\sigma^2 + \tau_\pi + \tau_\beta$ ) included in individual scores ( $\sigma^2$ ) was compared to the between group variance ( $\tau_\pi$ ) to estimate

the proportion of variance explained by differences in individual teachers' mean classroom scores, and between-school variance ( $\tau_\beta$ ) to evaluate the proportion of variability explained by school difference. Equations for proportioning within classroom (2), between-classroom (3), and between-school variances (4) are shown below.

$$\text{Within classroom} = \frac{\sigma^2}{(\sigma^2 + \tau_\pi + \tau_\beta)} \quad (2)$$

$$\text{Between-classroom} = \frac{\tau_\pi}{(\sigma^2 + \tau_\pi + \tau_\beta)} \quad (3)$$

$$\text{Between-school} = \frac{\tau_\beta}{(\sigma^2 + \tau_\pi + \tau_\beta)} \quad (4)$$

After partitioning the variance for each level of the model, conditional models were constructed to evaluate the significance of predictors at each level. Predictors were sequentially tested beginning with child level variables and following with classroom and school level predictors. The sequencing of variable entry was based on the hypothesis that most of the variability in teacher ratings will be explained by child level factors, followed by classroom and school characteristics. Level 1 predictors included child factor variables (e.g. gender, race/ethnic heritage, age at rating, special education status, disability accommodations status, gifted eligibility, and English Language Learning status). Equation 5 specifies that gender, race, age at rating in months, special education status, disability accommodations status, gifted status, and ELL services eligibility were added to the model to account for the variability in a teacher's ratings attributable to child characteristics.

#### Level 1

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk}(\text{gender})_{ijk} + \pi_{1jk}(\text{race})_{ijk} + \pi_{1jk}(\text{age})_{ijk} + \pi_{1jk}(\text{sped})_{ijk} + \pi_{1jk}(\text{dis})_{ijk} + \pi_{1jk}(\text{gifted})_{ijk} + \pi_{1jk}(\text{ELL})_{ijk} + \varepsilon_{ijk} \quad (5)$$

Thus,  $Y_{ijk}$  is the teacher rating for the  $i$ th student in the  $j$ th classroom in the  $k$ th school. Child gender, race, age, special education status, disability accommodations status and ELL eligibility were evaluated as predictors of teacher ratings of individual students. Additional level 1

analyses were conducted to determine if gender moderated the effects of racial/ethnic heritage and to estimate the contributions of three child factor variable groups: 1) race/ethnic heritage, 2) support services, and 3) age and gender.

The classroom mean scores were then adjusted in accordance with the characteristics of children enrolled in each classroom. Equation 6 states that after accounting for the characteristics of children in each classroom, teachers' classroom mean scores vary as a function of the grand mean of all classroom mean score ratings ( $\beta_{00k}$ ), and individual classroom characteristics ( $r_{0jk}$ ). In the Level 2 model, the percent of non-White non-Asian students and the total number of students in each classroom (e.g. class size) were added as classroom level predictors.

### Level 2

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}(\text{underserved})_{jk} + \beta_{02k}(\# \text{ students})_{jk} + r_{0jk} \quad (6)$$

$$\pi_{1jk} = \beta_{10k} + \beta_{11k}(\text{underserved})_{jk} + \beta_{12k}(\# \text{ students})_{jk} + r_{1jk}$$

Level 3 of the model included each of the school characteristics as predictors of teacher ratings of student scores (See Equation 7). For simplicity, school characteristics (average years of teaching experience, average percent of grade 3-5 students passing Smarter Balanced Assessment (SBA) (2015-2016) English Language Arts, average percent grade 3-5 students passing SBA Math (2015-2016), percent of students receiving lunch subsidy, and school enrollment) are represented in the model as a single variable; however, each of the school characteristics were evaluated independently in the Level 3 model.

### Level 3

$$\beta_{00k} = \gamma_{000} + \gamma_{000} (\text{school characteristics})_{00} + u_{00k} \quad (7)$$

$$\beta_{01k} = \gamma_{010} + \gamma_{010} (\text{school characteristics})_{00} + u_{01k}$$

$$\beta_{10k} = \gamma_{100} + \gamma_{100} (\text{school characteristics})_{00} + u_{10k}$$

$$\beta_{11k} = \gamma_{110} + \gamma_{110} (\text{school characteristics})_{00} + u_{11k}$$

Predictors from each level of the hierarchical model, were then added to a mixed model to evaluate the contributions of each variable in predicting teacher ratings and to identify cross-level interactions among child, classroom, and school level variables.

Following an analysis of the full sample, the analyses were repeated for Title 1 and Non-Title 1 school sub-samples to evaluate the relationship of predictors to teacher ratings of student social and emotional competency. After separate analyses for Title 1 and Non-Title 1 schools were conducted, interactions terms were created between child level predictors and school sector (Title 1 or Non-Title 1) to evaluate the impact of these cross-level interactions on teacher ratings of students' social and emotional competency.

Teacher responses to individual DESSA-Mini items were also examined to evaluate response patterns (e.g. same rating across all items, restricted range, rating extremes) within individual student ratings and across individuals in each classroom. In addition to the quantitative analysis for the full sample, a detailed analysis of teacher rating scale responses was conducted for the 15 teachers participating in cognitive interviews.

### **Qualitative Component Methods**

**Teacher recruitment and participants in semi-structured interviews.** All grade 3-5 certificated non-special education classroom teachers were included in the sampling frame (N = 173). To recruit teacher volunteers, the Director of Research and Accountability posted notice of

the study in weekly publication sent to all district administrators. Elementary principals were asked to forward an email written by the principal investigator to their grade 3-5 teachers and to encourage teacher participation. The Social and Emotional Learning Curriculum Developer sent a subsequent email to elementary instructional leaders asking them to encourage participation. Interested teachers were asked to express interest in an email the investigator.

Teachers were selected for interviews using a stratified purposive sampling plan. The initial stage of sampling clustered teachers based on the socioeconomic status (SES) of the building in which they teach. Elementary schools were clustered into Title 1 schools ( $n = 4$ ) and Non-Title 1 schools ( $n = 13$ ) based on the percentage of students receiving free or reduced price meals (“Title 1—Improving the Academic Achievement of the Disadvantaged” (1964)). In Title 1 schools, 40% or more of enrolled students are eligible for free or reduced price meals. Following initial Title 1, Non-Title 1 school clustering, schools with several volunteers were identified to allow for between teacher comparisons among educators working in the same school context. Individual teachers were then selected to optimize maximum variation based on teacher characteristics that included: gender, race/ethnicity, years of teaching experience, and classroom mean scores on the teacher rating scale.

Emails were sent to 15 grade 3-5 teachers to invite participation and schedule interview times. In advance of in-person interviews, participants were provided a Participant’s Agreement to review and sign. All participants held Washington state teacher certification. The years of teaching experience among participating teachers ranged from 1 year to 29 years ( $M = 11.13$  years). Thirteen females (87%) and 2 male teachers (13%) participated in interviews. The race and ethnicity of participating teachers reflected the racial and ethnic diversity among elementary

teaching staff within the district. Twelve participants identified as White (80%), two identified as Asian (13%), and one teacher identified as Hispanic/ Latina (6%).

The classroom mean SET t-scores of participating teachers (range T = 42.50-59.29, M = 52.93) were consistent with the classroom mean scores observed across the district (M = 53.31) (See Table 4.3). The average classroom Social and Emotional Total t-score means among participants was 52.93.

**Table 4.3.** *School SES, classroom means, and informant descriptors for interviewees*

Variables	Title 1	Non-Title 1
Schools	3	3
Teachers	6	9
Gender	5F	8F
Race/Ethnicity	5 White, 1 Latina	7 White, 2 Asian
Teaching Exp.	1-13 years (M = 8.67)	2-29 years (M = 12.78)
SET t-scores	42.50-59.29 (M=53.68)	43.70-54.72 (M=52.32)

**Procedure.** Teacher interviews took place in participating teachers’ classrooms, library study rooms, and coffee shops outside of school hours. Interviews commenced with informal conversation to establish rapport between the interviewer and participant. All interviews were audio-recorded. Field notes regarding researcher’s observations and participants’ responses were documented on an interview template (See template in Appendix B). Interviews lasted between 60 and 90 minutes and duration depended on the length of individual teacher’s responses and queries initiated by the researcher. Audiotaped interviews were transcribed into text documents using word processing software. Participating teachers received a \$100 MasterCard gift card to honor their time.

The semi-structured teacher interviews were conducted utilizing three cognitive interviewing techniques (e.g. think-aloud, verbal probing, expansive probing). See interview protocol in Appendix A. These techniques were selected to make teachers’ cognitive processes

explicit and visible (Advanced Research Seminar on Cognitive Aspects of Survey Methodology, Jabine, National Research Council (U.S.), & Committee on National Statistics, 1984; Beatty & Willis, 2007; Ginsburg, 1997; Miller, Chepp, Willson, & Padilla, 2014; Willis, 2005). Likert-type items were added to teacher interviews to probe teachers' comprehension of DESSA-Mini behaviors and the rater's level of confidence in rating the frequency of specific behaviors.

***Think-aloud.*** In the think-aloud portion of the interview, each teacher was asked to call to mind a student in the 'strength' range of the DESSA-Mini--'a student with strong social and emotional skills.' The participant was then asked to provide the student's gender and ethnicity. Teachers were also asked whether the child received supplementary services (e.g. special education, disability accommodations, English language learning, gifted curriculum). The participant was then presented with a printed copy of the DESSA-Mini and was instructed to *think-aloud* as she rated the student. Participants were encouraged to tell the researcher everything that came to mind as they rated individual students—'whether it seemed important or not' (Willis, 2005). The researcher was quiet during the *think-aloud* to minimize intrusions and facilitate the teacher's verbalization of thoughts (Beatty & Willis, 2007). Prompting for additional information was avoided. The *think-aloud* rating process was repeated as each teacher was asked to rate a student in the 'typical' range and a student in the 'need' range. Each participant rated three students. Sequencing the *think-aloud* activity at the beginning of the interview was intended to elicit teachers' thoughts without guiding participants to focus on specific aspects of the rating process. The *think-aloud* also primed teachers to answer targeted questions by reflecting on their recent ratings of students in their classroom (Beatty & Willis, 2007).

**Verbal probing.** Following the *think-aloud* portion of the interview, the researcher *verbally probed* (Willis, 2005) teachers' cognitive processes using questions designed to elicit teachers' thoughts about three DESSA-Mini items selected from Form 1 (e.g. pays attention, does something nice for somebody, contributes to group efforts). The three items were selected to address a range of social and emotional competency skills. For example, the 'pays attention' prompt was selected as a measure of the intrapersonal skill self-management and the 'contributes to group efforts' and 'does something nice for somebody' prompts were included as an indicators of students' interpersonal competencies.

*Verbal probe* questions prompted teachers to *define* behaviors included in DESSA-Mini items, to *recall* contexts in which the behaviors had been observed, to describe how they *formed judgments* about the frequency of students' skills, and to explain their thinking in *selecting a response option*. Anticipated probes (Beatty & Willis, 2007) were written (e.g. 'Tell me more, what do you mean', 'Say more about that') to prompt teachers to clarify, or elaborate on an unclear, or incomplete, response. At times, emergent probes were also used to prompt a teacher to expand her thinking, or explain her reasoning. These probes were flexible, unscripted, and spontaneously generated in reaction to teachers' responses (Beatty & Willis, 2007; Ginsburg, 1992). Likert-type response items were included in the *verbal probe* to assess the informant's comprehension of specific DESSA-Mini items and her level of confidence in rating the frequency of those behaviors for individual students.

**Expansive probing.** Following the *verbal-probe* of specific DESSA-Mini items, open-ended questions were asked to elicit expanded verbal descriptions of how teachers assess students' social and emotional skills (Ginsburg, 1997; Ginsburg, 1992). The use of open-ended questioning, or *expansive probing* (Beatty et. Al 1996, Willis, 2005), provided narrative

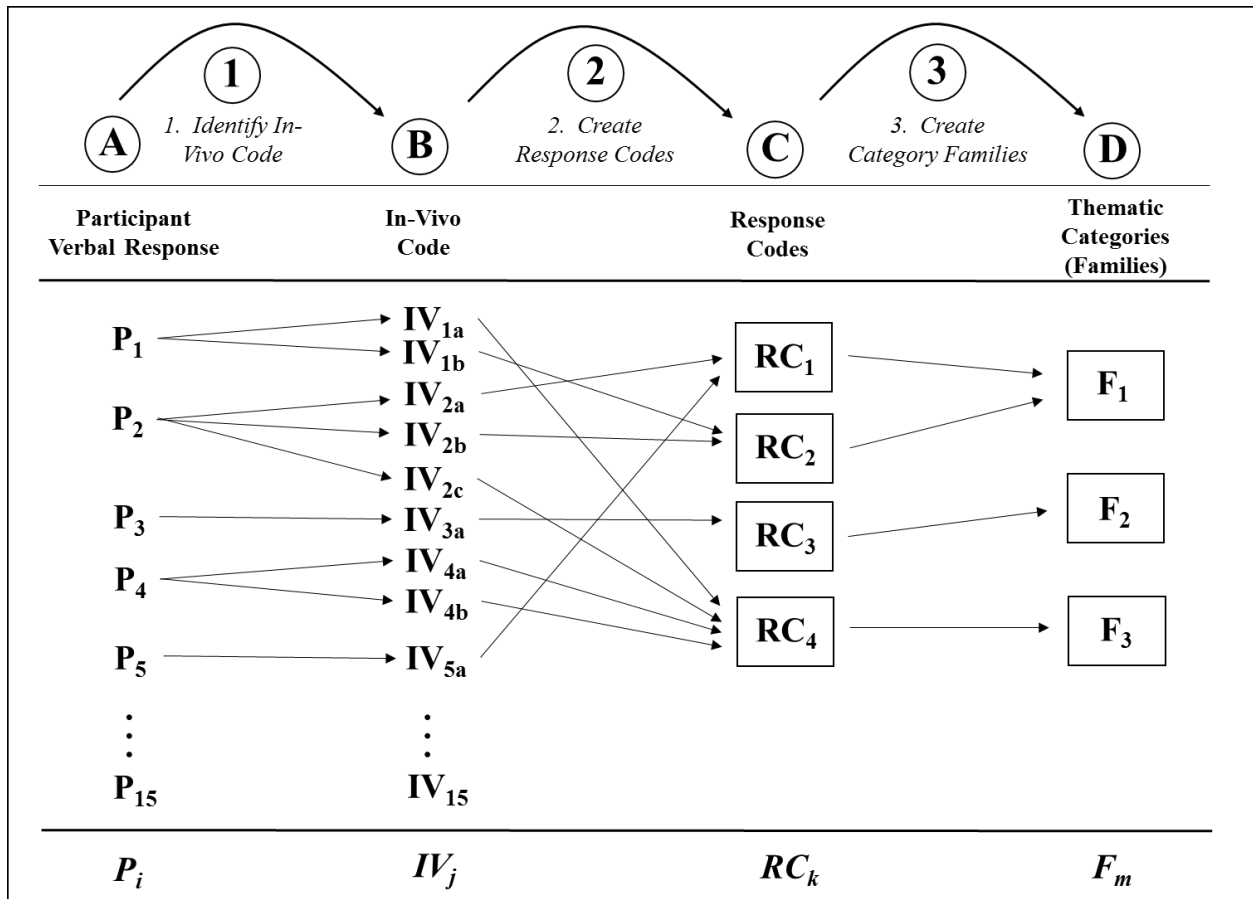
information about teachers' thoughts pertaining to the assessment of students' social and emotional skills. The *expansive probing* engaged teachers in reflecting on the thinking they do during their daily practices to identify factors independent of student behavior that may influence ratings. Teachers were asked how they assess the development of students' skills over time, and their thoughts pertaining the role culture and social context play in their evaluation of students' skills.

**Preparation of qualitative data for analysis.** Each transcribed interview was read in its entirety and open coding was used to identify themes (e.g. race/ethnicity, teacher-student relationship, teacher training), approaches to evaluation (e.g. top-down assessment, recollection of specific events), and responses to the measure itself (e.g. confusion, uncertainty, perspectives on appropriate usage). Additionally, notations were made regarding participants' references to factors influencing their ratings. Careful attention was paid to terms, phrases, or themes repeated by the participants.

Following an initial reading and open coding of each interview, in-vivo coding (Strauss, 1987; Corbin, 2015; Saldana, 2015) was used to identify short phrases or terms participants used in response to specific questions (e.g. what does it mean to pay attention?). In-vivo coding retained and highlighted the teacher voice and prevented an over application of the researcher's lens when interpreting interviews (Saldana, 2015). Interview segments were also coded for methodology (e.g. think-aloud, verbal, probe, expansive probe) and prompt (e.g. what does it mean to pay attention). These codes enabled a more detailed analysis of informants' responses and facilitated cross-informant analysis.

Research questions guided the selection of specific interview segments for analysis. Portions from each of the 15 interviews were tabled in data displays. Data displays included the

probe (e.g. what does it mean to pay attention?) and verbatim responses for each of the 15 participants. Words, phrases, or short selections from each participant’s interview emerged as in-vivo codes. The in-vivo codes were then sorted into similar themes, or ideas, and aggregated into response codes. The response codes that emerged from the in-vivo coding process were then used to re-code each of the interviews. Response codes were refined and aggregated into thematic categories. See Figure 4.1 for a visual summary of the coding process.



**Figure 4.1.** Coding process (adapted from Castro et al. model, 2010)

The iterative and recursive coding process was performed numerous times to complete the qualitative analysis.

A separate wave of coding, grounded in Cognitive Aspects of Survey Methodology tradition, was conducted to evaluate teachers’ thought processes when responding to the rating

scale. This phase of analysis focused specifically on the four cognitive enacted tasks (e.g. comprehension, retrieval, judgment, and response selection) when individuals respond to surveys, or rating scale type items (e.g. DESSA-Mini) (Advanced Research Seminar on Cognitive Aspects of Survey Methodology et al., 1984; Tourangeau & Rasinski, 1988). Coding during this phase was deductive and commenced with a broad application of thematic category codes (e.g. comprehension, retrieval, judgment, response selection). These theoretically informed thematic category codes were disaggregated into sub-codes to target specific segments of teacher interviews. The sub-codes were subsequently analyzed to identify specific themes. For example, the broad task code of ‘comprehension’ was segmented into 3 behavior codes (e.g. paying attention, contributing to group efforts, and doing something nice for somebody). Each of these behavior codes was subsequently disaggregated into specific themes contained within teacher interviews relevant to the behavior.

A third approach to coding was developed to investigate the ways in which teachers think about child, informant, and contextual factors when rating students social and emotional skills. Each of the 15 interviews were read to identify instances in which the teacher referred to child, classroom, school, or informant characteristics when rating students or sharing their thoughts about students’ social and emotional development. The selections of teacher interviews were coded with broad thematic family codes (e.g. child factors, context factors, informant factors). Broad literature informed codes were then disaggregated into response codes to facilitate a more detailed analysis of how teachers think about these factors when evaluating students’ social and emotional skills. The ‘child factors’ code, for example, was disaggregated into 5 response code subcategories: a) child cultural factors; b) child family factors; c) child academic factors; d) child temperament factors; and e) child gender factors.

Rigorous coding facilitated a detailed analysis of informants' responses to locate similarities and differences within individual teacher responses, and across respondents. Additionally, precise codes supported the identification of inter-code patterns that helped to illustrate the ways in which teachers think about children's social and emotional skills. Following coding by the primary investigator, a subset of the data was coded by a doctoral level independent researcher with expertise in the field of social and emotional learning. The researcher was trained in the coding schemes utilized in the analysis and was provided with example interview segments to code. The rater randomly selected interview segments containing behavioral definition codes, frequency codes, and measures of congruence from the participant interviews. The random selection process based on code segmented supported analysis over a greater number of participant interviews. Teacher response styles were also coded by the independent rater. The overall inter-rater reliability was 94.75 % agreement. Agreement on the frequency codes and response styles was 100%. Agreement on measures of congruence was 89% and for definitional codes, the agreement rate was 90%.

**Data analysis.** Codes (e.g. open codes, in-vivo codes, response codes, and thematic category codes) were entered in Atlas.ti 7 for Mac to facilitate qualitative analysis. Each code was analyzed to determine if the level of specificity supported the analysis. Excessively broad codes were disaggregated to increase specificity. Unique and low frequency codes were aggregated, eliminated, or retained based on the theoretical importance and/or utility of the code in understanding teacher thoughts pertaining to the assessment of students' social and emotional skills. Analytic memos documented preliminary findings and developed hypotheses during the data preparation process. A final list of codes and code definitions is included in Appendix D.

Findings derived from the quantitative investigation guided the analysis of qualitative data. In addition to describing broad patterns in teachers' thought processes and ratings, I attended to interview segments that helped to explain quantitative findings. The analysis explored the ways in which teachers described students, and teachers' recall of events, tendencies, and specific child behaviors offered in support of their ratings. The interviews were also mined for teacher references to child, classroom, and school level characteristics influencing their ratings.

The congruence between individual teacher's thoughts and their related practices were analyzed. Each rater's definition of specific behaviors and her assessments of the frequency with which students display those behaviors was evaluated by comparing the rater's definition to her *think-aloud* ratings of the same behavior. Ratings for each of the three students were evaluated to assess the congruence between each teacher's definitions and their assessment practices. The analysis investigated whether elements included in teacher definitions were specifically mentioned, or referenced, during their ratings of individual students. Each student rating (e.g., S<sub>1</sub>, S<sub>2</sub>, S<sub>3</sub>) was coded with an 'M' for specific mention of behavior(s) contained within the teacher's definition of the behavior, or an 'R' for behavior referenced in the definition, for each of the three behaviors analyzed (e.g. paying attention, contributing to group efforts, doing something nice for somebody). For example, if an informant defined paying attention as a set of physical behaviors that suggest a student is attending (e.g. eyes on the speaker, nodding), and explicitly mentioned physical behaviors (e.g. eyes on the speaker, nodding) when rating students, that rating of attending behavior was coded with a 'M' to indicate that elements from the teacher's definition were mentioned in her think-aloud rating for that student. Ratings in which an element of the definition was suggested, but not specifically mentioned, were coded with a

‘R’ to indicate that a behavior was referenced by the informant. A rating that referenced doodling or talking to a friend, for example, does not explicitly mention physical behaviors suggestive of attention (e.g. eyes on speaker, nodding); however, the rating referenced physical acts that may be associated with not attending and so the rating was coded with an ‘R.’

The congruence codes applied to the three teachers’ ratings for each behavior were then converted to a numerical score using a 5-point rubric (see Table 4.4).

**Table 4.4.** *Congruence rubric for evaluating teacher definitions and ratings*

Score	Description
0	Elements are not referenced AND are not mentioned in any of the three ratings
1	Elements referenced in up to two ratings AND are not mentioned in any of three ratings
2	Elements are referenced in 3 ratings OR mentioned in at least 1 rating
3	Elements mentioned in 2 ratings
4	Elements mentioned in all 3 ratings

Separate congruence scores were calculated for each of the three probed behaviors (e.g., attention =  $C_A$ , something nice,  $C_N$ , group efforts =  $C_G$ ). Rubric scores for each rater (e.g.  $C_A$ ,  $C_N$ ,  $C_G$ ) were then averaged to provide an overall measure of internal congruence ( $C_T$ ) for each teacher.

Teachers responses to questions about their comprehension of DESSA-Mini items and their confidence in their ability to accurately their students on frequency which they exhibit these behaviors were also evaluated using 5-point Likert-type scales (e.g., 0 = No comprehension, 1= Uncertain of comprehension, 2 = Somewhat comprehend, 3 = Comprehend, 4 = Fully Comprehend). Teacher averages were calculated to describe each teacher’s comprehension of

the rating scale item and her confidence in using the measure to accurately rate the frequency of probed behaviors.

## Chapter 5: Results and Discussion

### Quantitative Findings

*Item level analysis.* Zero-order correlations were calculated to evaluate the relationships among the eight DESSA-Mini items. The inter-item correlations are shown in Table 5.1. Each of the eight items are positively correlated with the other items. The strength of the relationships range from  $r = 0.566$  to  $r = 0.793$ .

**Table 5.1.** *Inter-item correlations among DESSA-Mini Items*

<i>Item</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>
<sup>a</sup> Responsibility	1	.657**	.619**	.699**	.627**	.645**	.656**	.731**
<sup>b</sup> Nice	.657**	1	.733**	.583**	.633**	.571**	.599**	.664**
<sup>c</sup> Speak	.619**	.733**	1	.560**	.632**	.568**	.566**	.633**
<sup>d</sup> Attention	.669**	.583**	.560**	1	.720**	.793**	.738**	.638**
<sup>e</sup> Group	.627**	.633**	.632**	.720**	1	.735**	.676**	.622**
<sup>f</sup> Tasks in order	.645**	.571**	.568**	.793**	.735**	1	.771**	.650**
<sup>g</sup> Shows care	.656**	.599**	.566**	.738**	.676**	.771**	1	.690**
<sup>h</sup> Follows advice	.731**	.664**	.633**	.638**	.622**	.650**	.690**	1

\*\*Correlation is significant at the 0.01 level (2-tailed)

Descriptive statistics for teacher responses to each DESSA-Mini item (shown in Table 5.2.) were calculated to evaluate mean score differences among items. Item mean scores were comparable overall with a slightly lower mean score ( $M = 2.80$ ) observed for ‘pays attention’ and a slightly higher score observed for ‘follows advice of trusted adult’ ( $M = 3.14$ ).

**Table 5.2.** *Descriptive statistics for individual DESSA-Mini items*

<i>Item</i>	<i>N</i>	<i>Mean</i>	<i>SD</i>	<i>Variance</i>
Responsibility	4005	2.97	.932	.869
Nice	4005	2.88	.919	.845
Speak	4005	2.92	.876	.768
Attention	4005	2.80	.977	.954
Group	4005	2.84	.955	.912
Tasks in order	4005	2.93	.920	.846
Shows care	4005	3.00	.906	.822
Follows advice	4005	3.14	.790	.625

The frequency of teacher responses to individual items, and the variability in teacher responses within ratings of individual students were also evaluated to provide greater detail on informant response patterns. Item response frequencies are shown in Table 5.3. Teacher responses for each of the 8 items were distributed across the full range of available responses. Accordingly, the range of responses for each of the 8 items spanned from a minimum score of zero (0 = never) to maximum score of four (4 = very frequently). Most teacher ratings for each of the items favorably described students as ‘frequently’ or ‘very frequently’ exhibiting the skill.

**Table 5.3.** *Response frequencies for individual DESSA-Mini items*

<i>Item</i>	<i>Number of ratings per response option</i>				
	<i>0</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
Responsibility	52	249	748	1686	1270
Nice	38	253	980	1626	1108
Speak	36	221	830	1855	1063
Attention	51	369	972	1543	1070
Group	53	321	924	1635	1072
Tasks in order	41	282	745	1774	1163
Shows care	25	249	748	1662	1321
Follows advice	14	113	590	1879	1409

Rating descriptors: 0 = Never, 1 = Rarely, 2 = Occasionally, 3 = Frequently, 4 = Very Frequently

The between item variance in teacher ratings within individual student ratings (shown in Table 5.4) was evaluated to investigate patterns in teachers’ responses. Student ratings containing no variability in teacher responses across items (e.g. all items rated 3) comprised 23.3% of the total sample. Teacher ratings of students in which one item differed from the remaining items by a single point in rating comprised 13.6% of the total sample.

**Table 5.4.** *Variance within-student ratings*

<i>Variance</i>	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative Percent</i>
.000	934	23.3	23.3
.125	546	13.6	37.0
.214	528	13.2	50.1
.268	535	13.4	63.5
.411	352	08.8	72.3
.500	209	05.2	77.5
.554	154	03.8	81.3
.696	172	2.2	85.6

The frequency of inter-item variance within individual ratings was also evaluated at the classroom and school levels. At the classroom level, the number of student ratings with no variability across items ranged from a minimum of one student (min =1) to a maximum of 23 students (max =23). At the school level, the number of student ratings with no variability in response choices across items ranged from a minimum = 8.82% of total student ratings to a maximum = 37.69% of total ratings. The variability in teacher responses across items for individual students was greater at schools with larger percentages of students eligible for free or reduced meals. The variance across items within individual ratings was positively correlated with the percentage of free and reduced lunch students enrolled in a school ( $r = .592, p < 0.05$ ), and statistically significant at the 0.05 level in a two-tailed test.

**Descriptive analyses.** Means and standard deviations for DESSA-Mini teacher ratings of student social and emotional competency for the total sample are shown in Table 5.5. Group mean social emotional total (SET) scores were generated for student gender, race/ethnic heritage, and eligibility for the following services: disability accommodations, special education, gifted programs, and English language learning. Means and standard deviations were also calculated to provide grade level mean scores.

**Table 5.5.** *SET Means and standard deviations for child factors for total sample*

<i>Group</i>	<i>Social and Emotional Total</i>		
	<i>N</i>	<i>Mean</i>	<i>SD</i>
ISFemale	1956	56.41	10.25
NoFemale	2049	50.47	10.38
ISELL	423	48.39	10.13
NoELL	3582	53.96	10.94
ISSped	248	44.26	10.40
NoSped	3757	53.97	10.48
ISDis	112	48.04	9.24
NoDis	3893	53.53	10.74
ISGifted	600	54.12	9.44
NoGifted	3405	53.24	10.94
ISAsian	1511	54.28	10.33
ISBlack/African-American	100	47.91	10.91
ISHispanic/Latino	483	50.41	10.84
ISMulti-Ethnic	332	53.45	10.78
ISWhite	1548	53.76	10.82
Grade3	1394	52.29	10.70
Grade4	1375	53.09	10.55
Grade5	1236	54.90	10.81
Total	4005	53.31	5.01

Note: ISFemale = student identified as female, NoFemale = student identified as male, ISELL = student receives English language learning services, NoELL = student is not eligible for English language services, ISSPED = student receives special education services, NoSped = student is not eligible for special education services, ISDis = student is eligible for accommodations based on disability, NoDis = student is not eligible for accommodations based on disability, ISGifted = student identified as intellectually gifted, NoGifted = student not identified as intellectually gifted, IsAsian = student identified as Asian, ISBlack/African-American = student identified as Black/African-American, IsHispanic/Latino = student identified as Hispanic or Latino, IsMulti-Ethnic = student identified as belonging to two or more racial/ethnic groups, IsWhite = student identified as White or Caucasian.

*Title 1 and Non-Title 1 school descriptive statistics.* Separate descriptive analyses were generated to evaluate SET mean scores and standard deviations for groups of students enrolled in Title 1 and Non-Title 1 schools. SET means scores and standard deviations for child factors among students attending Title 1 schools are shown in Table 5.6. The number of students ( $n = 577$ ) attending Title 1 schools ( $n = 4$ ) represents 14.41% of the total sample ( $N = 4005$ ).

**Table 5.6.** SET Means and standard deviations for child factors at Title 1 Schools

<i>Group</i>	<i>Social and Emotional Total</i>		
	<i>N</i>	<i>Mean</i>	<i>SD</i>
ISFemale	275	53.60	10.54
NoFemale	302	46.25	10.28
ISELL	181	46.00	9.57
NoELL	396	51.47	11.23
ISSped	48	41.23	9.26
NoSped	529	50.53	10.85
ISDis	27	45.26	9.46
NoDis	550	49.98	11.05
ISGifted	0	-----	-----
NoGifted	577	49.76	11.02
ISAsian	121	52.97	11.20
ISBlack/African-American	46	46.48	9.12
ISHispanic/Latino	213	48.25	10.45
ISMulti-Ethnic	35	50.57	10.79
ISWhite	157	50.35	11.76
Grade3	182	50.37	11.84
Grade4	216	48.13	10.69
Grade5	179	51.09	10.35
Total	577	49.76	11.02

Descriptive statistics for children ( $n = 3428$ ) attending Non-Title 1 schools ( $n = 13$ ) are shown in

Table 5.7. The SET Means for each of the groups is higher among students enrolled in Non-

Title 1 schools.

**Table 5.7.** *SET Means and standard deviations for child factors at Non-Title I Schools*

<i>Group</i>	<i>Social and Emotional Total</i>		
	<i>N</i>	<i>Mean</i>	<i>SD</i>
ISFemale	1681	56.87	10.13
NoFemale	1747	51.20	10.23
ISELL	242	50.18	10.19
NoELL	3186	54.27	10.54
ISSped	200	44.99	10.54
NoSped	3228	54.54	10.31
ISDis	85	48.92	9.05
NoDis	3343	54.54	10.31
ISGifted	600	54.12	9.44
NoGifted	2828	53.95	10.79
ISAsian	1390	54.40	10.25
ISBlack/African-American	54	49.13	12.20
ISHispanic/Latino	270	52.11	10.86
ISMulti-Ethnic	297	53.85	10.73
ISWhite	1391	54.15	10.52
Grade3	1212	52.58	10.49
Grade4	1159	54.02	10.26
Grade5	1057	55.54	10.56
Total	3428	53.98	10.56

*Zero order correlations.* Zero-order correlations were calculated to evaluate the independence of predictors at each level of the model. The Pearson correlation coefficients for child level variables are shown in Table 5.8. As expected, race/ethnic predictors were collinear with other race/ethnic factors; however, the correlation coefficients among remaining predictors fell below the  $r = 0.90$  threshold (Tabachnick, & Fidell, 2012) suggestive of multicollinearity, and therefore do not violate collinearity assumptions.

**Table 5.8.** *Correlations among student level variables for total sample*

<i>Variable</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>	<i>I</i>	<i>j</i>	<i>k</i>
<sup>a</sup> Age	1	.095**	.046**	.033*	-.051**	.085**	.019	.022	.028	.023	.021
<sup>b</sup> SET	.095**	1	-.218**	-.084**	-.159**	.029	.001	-.014	-.005	-.009	-.021
<sup>c</sup> IsSPED	.046**	-.218	1	-.044**	.070**	-.073**	-.030	.014	-.005	-.018	-.009
<sup>d</sup> ISDis	.033*	-.084	-.044**	1	-.024	-.041	-.023	-.010	.005	-.011	-.010
<sup>e</sup> IsELL	-.051**	-.159*	.070**	-.024	1	-.140**	.022	.021	-.005	.012	.053**
<sup>f</sup> IsGifted	.085**	.029	.073**	-.041**	.140**	1	.026	-.028	-.055**	-.027	-.039*
<sup>g</sup> IsAsian	.019	.001	-.030	-.023	.022	.026	1	.984**	.957**	.979**	.975**
<sup>h</sup> IsBlack	.022	-.014	-.014	-.010	.021	-.028	.984**	1	.984**	.994**	.992**
<sup>i</sup> IsWhite	.028	-.005	.005	.005	-.005	-.055**	.957**	.984**	1	.979**	.975**
<sup>j</sup> IsMulti	.023	-.009	-.011	-.011	-.012	-.027	.979**	.994**	.979**	1	.988**
<sup>k</sup> IsLatino	.021	-.021	-.009	-.010	.053**	-.039**	.975**	.992**	.975**	.988**	1

\*\*Correlation is significant at the 0.01 level (2-tailed)

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

The classroom level covariates--the percentage of underserved students and class size--are negatively correlated. The relationship is statistically significant in a two-tailed test (shown in Table 5.8), but small ( $r = -.210, p < 0.01$ ). Although the correlation is significant, the strength of the relationship small and not suggestive of collinearity. The observed relationship is consistent with the participating district's funding policies mandating lower class sizes in schools serving higher percentages of students receiving free or reduced meals. Title-1 schools also serve a larger proportion of students from non-White and non-Asian racial and ethnic groups.

**Table 5.9.** *Correlations among classroom level variables for total sample*

<i>Variable</i>	<i>a</i>	<i>b</i>
<sup>a</sup> Class Size	1	-0.210**
<sup>b</sup> P_Disadvantaged	-0.210**	1

\*\*Correlation is significant at the 0.01 level (2-tailed)

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

Note: Class Size = number of students enrolled in class at time of rating, P\_Disadvantaged = % of students from historically disadvantaged racial/ethnic groups (non-Asian and non-White)

At the school level, multicollinearity was observed between the mean percentage of grade 3-5 students passing the state English Language Arts (ELA) achievement measure and the mean percentage of students passing the grade 3-5 math achievement measure (MTH).

Multicollinearity was also observed between each of the state academic achievement measures

and the percentage of students receiving free or reduced lunch (See Table 5.8). Accordingly, the academic achievement variables were eliminated as school level predictors from subsequent multilevel models.

**Table 5.10.** *Correlations among school level variables for total sample*

<i>Variable</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>
<sup>a</sup> AvgTeach	1	-.196	.422	.219	.266
<sup>b</sup> P_Free	-.196	1	-.980**	-.938**	-.301
<sup>c</sup> MP_ELA	.422	-.980**	1	.956**	.590**
<sup>d</sup> MP_MTH	.219	-.938**	.956**	1	.633**
<sup>e</sup> Enroll	.302	.241	.016	.009	1

\*\* Correlation is significant at the 0.01 level (2-tailed)

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

**HLM Analyses.** Hierarchical linear modeling (HLM) was used to analyze predictors in teacher ratings of students’ social and emotional competency. The data was structured to reflect the nesting of students (level 1) within classrooms (level 2), and classrooms within schools (level 3). The analyses were conducted in several waves that included the construction of a fully unconditional model, means as outcomes models, random coefficient models, and intercepts and slopes as outcome models. (Raudenbush, 2002). The full sample was also disaggregated into Title 1 and Non-Title 1 sub-samples to investigate whether school contexts moderated the relationship between predictors and teacher ratings of students’ social and emotional competency. Descriptive statistics for the full sample, Title 1 sub-sample and Non-Title 1 sub-sample are shown in Tables 5.11, 5.12, and 5.13, respectively.

**Table 5.11.** *Descriptive statistics for variables at each level of model for total sample*

<i>Variable</i>	<i>N</i>	<i>Mean</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
Level 1					
SET Score	4005	53.37	0.81	28.00	71.00
Level 2					
M_SET Score	173	53.31	5.01	42.50	68.00
Class Size	173	23.15	2.97	10.00	30.00
P_ISFemale	173	48.93	7.62	31.60	68.20
P_ISSPED	173	6.20	5.42	00.00	25.00
P_ISDIS	173	2.84	3.84	00.00	19.00
P_ISELL	173	10.89	11.71	00.00	61.90
P_ISGIFTED	173	14.29	32.44	00.00	100.00
P_DISADV	173	24.34	16.64	00.00	90.48
Level 3					
GM_SET Score	17	52.83	2.48	48.22	56.68
GP_ISSPED	17	6.79	3.11	1.14	13.08
GP_ISDIS	17	2.99	1.85	00.68	7.66
GP_ISELL	17	13.21	11.73	01.82	40.88
GP_ISGIFTED	17	10.18	17.96	00.00	50.84
GP_DISADV	17	28.86	15.83	10.95	56.80
GP_AVGYTEACH	17	9.71	1.72	6.30	12.70
GP_FREE	17	19.26	17.99	1.90	54.30
GP_ENROL	17	522.82	120.47	257.00	731.00

Note: GP\_AVGYTEACH = the average number of years of teaching experience amongst teachers in school, GP\_Free = percent of children in school eligible for meal subsidies, GP\_Enrol = the number of students enrolled in October 2015

**Table 5.12.** Descriptive statistics for variables at each level of model for Title 1 schools

<i>Variable</i>	<i>N</i>	<i>Mean</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
Level 1					
SET Score	577	49.76	11.02	28.00	71.00
Level 2					
M_SET Score	27	49.75	4.72	42.50	59.20
Class Size	27	21.37	1.96	14.00	25.00
P_ISFemale	27	47.94	6.85	36.40	59.10
P_ISSPED	27	8.19	5.20	0.00	18.20
P_ISDIS	27	4.64	4.96	0.00	19.00
P_ISELL	27	31.44	11.98	14.30	61.90
P_ISGIFTED	27	-----	-----	-----	-----
P_DISADV	27	52.01	12.88	31.82	90.48
Level 3					
GM_SET Score	4	50.00	2.02	48.22	52.10
GP_ISSPED	4	8.04	2.59	4.75	10.61
GP_ISDIS	4	4.62	2.45	1.66	7.66
GP_ISELL	4	31.88	8.03	21.60	40.88
GP_ISGIFTED	4	-----	-----	-----	-----
GP_DISADV	4	52.47	4.23	47.45	56.80
GP_AVGYTEACH	4	9.15	2.31	6.30	11.80
GP_FREE	4	47.75	4.63	43.90	54.30
GP_ENROL	4	455.00	57.32	391.00	511.00

**Table 5.13.** *Descriptive statistics for variables at each level of model for Non-Title 1 schools*

<i>Variable</i>	<i>N</i>	<i>Mean</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
Level 1					
SET Score	3428	53.98	10.56	28.00	71.00
Level 2					
M_SET Score	146	53.97	4.79	43.70	68.00
Class Size	146	23.48	3.01	10.00	30.00
P_ISFemale	146	49.11	7.76	31.60	68.20
P_ISSPED	146	5.83	5.39	0.00	25.00
P_ISDIS	146	2.51	3.52	0.00	16.00
P_ISELL	146	7.09	6.61	0.00	21.70
P_ISGIFTED	146	16.93	34.69	0.00	100.00
P_DISADV	146	19.22	11.40	0.00	69.23
Level 3					
GM_SET Score	13	53.69	1.92	49.77	56.68
GP_ISSPED	13	6.40	3.25	1.14	13.08
GP_ISDIS	13	2.49	1.39	0.68	5.83
GP_ISELL	13	7.47	3.91	1.82	14.73
GP_ISGIFTED	13	13.31	19.63	0.00	50.84
GP_DISADV	13	21.60	9.30	10.95	38.55
GP_AVGYTEACH	13	9.88	1.57	6.30	12.70
GP_FREE	13	10.50	8.52	1.90	28.90
GP_ENROL	13	543.69	128.55	257.00	731.00

**Baseline models and intra-class correlations.** Fully unconditional models, were constructed for each of the three data sets to partition the variance explained by each level of the model. The variance components for the full sample, Title 1 schools sub-sample, and Non-Title schools sub-sample are shown in Table 5.14, Table 5.15, and 5.16, respectively.

**Table 5.14.** *Partitioning of variance for total sample*

	<i>Social and Emotional Total</i>
Within-class	81.79**
Between-class	15.60**
Between-school	2.61**

\*\*  $p < .01$ , \*  $p < .05$  on  $\chi^2$  tests of the variance components

In the full sample, the major source of variability in teacher ratings of students social and emotional competency was found at the within classroom or student level, however, a sizable

portion of variance (18.21%) was collectively located at the classroom (15.60%) and school levels (2.60%). The intercept only model revealed an intra-class correlation (ICC) of .1560 at the classroom level and .0261 at the school level. Accordingly, approximately 15.60% of variance in teacher ratings of student social and emotional competency is attributable to between classroom, or teacher differences, and 2.61% is associated with between-school level differences.

The intra-class correlation (ICC) calculated for the sub-sample containing Title 1 schools places an even greater proportion of the total variability in scores at the within classroom level. In contrast to the proportion of variance explained by within-class differences in the full sample (81.79%), an even greater proportion of the total variability in student scores is explained by within-classroom differences in Title-1 schools. The variability explained by within-classroom differences in Title 1 schools is 86.44% which is an increase of 4.79% of the total variance located at the within classroom level. The between-classroom, or between teacher, variability (13.31%) is lower for teachers teaching in Title 1 schools than among teachers across the total sample (15.6%). Relatively little variability exists at the between-school level among the four Title 1 schools included in the sample. In segmenting the full sample, based on Title 1 status, we find that between-school variance accounts for only 0.25% of the total variability in student scores in comparison to the 2.61% across the total sample. Teacher ratings within schools serving a similar demographic of student were more alike than teacher ratings between-schools serving a broader spectrum of students.

**Table 5.15.** *Partitioning of variance for Title 1 schools*

	<i>Social and Emotional Total</i>
Within-class	86.44**
Between-class	13.31**
Between-school	0.25**

\*\*  $p < .01$ , \*  $p < .05$  on  $\chi^2$  tests of the variance components

A similar shift in the proportion of variance explained by each level of the nested model was observed in the intra-class correlation for Non-Title 1 schools (see Table 5.16). Among Non-Title schools, the proportion of total variance explained by within-class differences increased to 83.29% from 81.79% in the full sample. The variation in outcome scores attributable to differences between-classrooms, or between teachers, teaching in Non-Title 1 schools increased from 15.60% observed in the total sample to 16.60% in the Non-Title 1 subsample. The variance attributable to between-school differences remains statistically significant, but quite small with only 0.11% of total variability in student scores explained by between-school factors.

**Table 5.16.** *Partitioning of variance for Non-Title 1 schools*

	<i>Social and Emotional Total</i>
Within-class	83.29**
Between-class	16.60**
Between-school	0.11**

\*\*  $p < .01$ , \*  $p < .05$  on  $\chi^2$  tests of the variance components

**Conditional models.** In the second wave of analysis, conditional baseline models were constructed to evaluate the contributions of predictor variables at each level of the model. The level 1, or child predictors (e.g. age-at-rating, race/ethnic heritage, gender, ELL status, Sped status, disability accommodations, and gifted eligibility), were imputed into a random-regression coefficients model (Raudenbush, 2002). The age-at-rating covariate was entered as a fixed variable centered on the group mean and the remaining demographic and service eligibility factors were entered as un-centered fixed variables. White served as the reference group for racial and ethnic analysis. Results are shown in Table 5.17.

**Table 5.17.** *HLM level 1 base model by social emotional total for full sample*

<i>Student Effects</i>	Social and Emotional Total		
	<i>Coefficient</i>	<i>se</i>	<i>t-ratio</i>
Intercept	51.73	0.50	102.57**
AgeatRating	0.14	0.03	4.03**
ISFemale	5.46	0.29	19.04**
ISSped	-8.92	0.60	-14.85**
ISDis	-5.39	0.87	-6.18**
ISELL	-3.62	0.51	-0.71**
ISGifted	0.25	0.80	0.32
ISAsian	0.72	0.36	2.04*
ISBlack/African-American	-4.58	0.94	-4.87**
ISMultiEthnic	-0.51	0.55	-0.92
ISLatino/Hispanic	-1.28	0.51	-2.50*

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

Note: AgeatRating = student's age in months at time of rating

Among the child factors, a student's gender, age at the time of rating relative to his classmates, and eligibility for special education services, English language services, and/or disability accommodations significantly predicted teacher ratings of student social and emotional competency. In comparison to male students, female students, on average, scored 5.46 points higher in social and emotional competency than their male peers ( $t = 19.04$ ,  $p = < 0.001$ ). This finding conforms to previous literature documenting gender differences in the development of social and emotional skills among school age children (Naglieri et al., 2009). Older students within classrooms were also predicted to earn higher teacher ratings than their younger classmates. For every additional month in age beyond the classroom average, students were expected to earn a 0.14 ( $t = 4.031$ ,  $p < 0.001$ ) increase in their SET t-score.

Significant differences in teacher ratings were also observed among groups of students receiving additional school services. Eligibility for special education services, on average, predicted an 8.92-point decrease in social and emotional competency score ( $t = -14.85$ ,  $p <$

0.001). Reductions in predicted score were also observed among students receiving disability accommodations (-5.39) ( $t = -6.17, p < 0.001$ ) and ELL services (-3.62) ( $t = -7.12, p < 0.001$ ). In contrast to the other service eligibility predictors, eligibility for gifted services was not a significant predictor of teacher ratings of social and emotional competency ( $t = 0.320, p > 0.50$ ).

Among the racial and ethnic predictors, being Asian, Black/African-American or Latino/Hispanic (in comparison to the White reference group) significantly predicted teacher ratings of social and emotional competency. On average, Asian students were predicted to have a .72-point increase in their SET t-score in comparison to their White peers ( $t = 2.04, p < 0.05$ ), Black/African-American students were predicted to score 4.58 points lower than their White classmates ( $t = -4.87, p < 0.001$ ), and the effect of being Latino predicted a 1.28-point decrease in score ( $t = -2.50, p < 0.05$ ). In contrast, being identified Multi-Ethnic did not significantly predict teacher ratings of SET ( $t = -0.92, p > 0.05$ ).

In combination, the level 1 child factor covariates explained an additional 17.24% of the within classroom variance [proportion variance explained =  $(\sigma^2 \text{ empty model} - \sigma^2 \text{ level 1 base model}) / (\sigma^2 \text{ empty model})$ ] or  $(94.66 - 78.34) / 94.66 = 17.24$ ) when compared to the fully unconditional model. Thus, student level predictors (e.g. age, gender, race/ethnic heritage, and eligibility for sped, disability accommodations, and ELL services) reduced the within-classroom variability in student scores by 17.24%.

*Race-gender interactions.* Interaction terms between race and gender were created to determine if student gender moderated the effect of belonging to racial/ ethnic groups on teacher ratings of students' social and emotional skills. Interaction terms were created for Asian females, Latino females, Black/African-American females, and Multi-Ethnic females. These interaction terms were added with race and gender factors to the baseline model. None of the

interaction variables significantly predicted teacher ratings of students' social and emotional competency indicating that gender does not moderate the relationship between a child's race/ethnic heritage and its effect on teacher ratings of student social and emotional competency. See table 5.18 for results of level 1 base model with interaction terms.

**Table 5.18.** *HLM level 1 base model by social emotional total with interactions*

<i>Student Effects</i>	<i>Social and Emotional Total</i>		
	<i>Coefficient</i>	<i>se</i>	<i>t-ratio</i>
Intercept	50.51	0.58	87.43**
ISFemale	5.71	0.47	12.06**
ISAsian	0.53	0.49	1.08
ISBlack/African-American	-4.25	1.43	-2.97**
ISMultiEthnic	-0.70	0.79	-0.88
ISLatino/Hispanic	-2.04	0.70	-2.92**
Female x Asian	0.59	0.67	0.87
Female x Black/African-American	-0.94	1.94	-0.48
Female x MultiEthnic	1.42	1.13	1.26
Female x Latino/ Hispanic	-0.38	0.98	-0.39

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

*Title 1 and Non-Title 1 baseline models.* Separate level 1 predictor models were constructed for the Title 1 and Non-Title 1 school data sets to evaluate if the effects of child level vary with school context (see Table 5.19). Within Title 1 Schools, a student's age at the time of rating (relative to classroom peers) is not a significant predictor of teacher ratings when controlling for other student factors. In contrast to the significant effects of being Latino or eligible for disability accommodations found within the full sample of schools, eligibility to disability accommodations and Latino/ Hispanic ethnic heritage were not found to be significant predictors of teacher ratings among students in Title 1 schools. The relatively large percentage of Latino/Hispanic students enrolled in Title 1 schools (36.92% of total enrollment) in comparison to the representation of Latino/Hispanic students within full sample (12.06 % of total enrollment) may minimize perceived differences between Latino students and others. Latino

students, comprising 36.92% of the student population within Title 1 schools are likely compared to other Latino students, thereby minimizing distinctions that may occur when the minority status of an underrepresented group is accentuated because of their level of representation in comparison to a sizable majority reference group (for discussion of reference group effects see Heine, Lehman, Peng, & Greenholtz, 2002).

In Non-Title 1 schools, being identified as Asian did not significantly predict teacher ratings of students' social and emotional competency after holding classroom and school membership constant. The proportion total students identified as Asian in the Non-Title 1 schools ( $n = 1390$ , 40.55%) is only slightly higher than the proportion of Asian students within the full sample ( $n = 1511$ , 37.73%).

**Table 5.19.** *HLM level 1 base model by social emotional total for Title 1*

<i>Student Effects</i>	Social and Emotional Total						
	<i>Coef.</i>	Title 1 Schools			Non-Title 1 Schools		
		<i>se</i>	<i>t-ratio</i>		<i>Coef.</i>	<i>se</i>	<i>t-ratio</i>
Intercept	48.37	1.30	37.10**	52.35	0.48	108.25	
AgeatRating	0.12	0.10	1.20	0.14	0.37	3.83**	
ISFemale	6.16	0.76	8.14**	5.33	0.31	17.22**	
ISSped	-7.71	1.36	-5.66**	-9.20	0.67	-13.78**	
ISDis	-2.93	1.80	-1.62	-5.94	1.00	-5.95**	
ISELL	-4.41	0.91	-4.85**	-3.20	0.62	-5.14**	
ISAsian	3.84	1.13	3.84**	0.37	0.37	1.00	
ISBlack	-3.45	1.50	-2.30*	-4.38	1.25	-3.51**	
ISMultiEthnic	-0.74	1.68	-0.44	-0.50	0.58	-0.87	
ISLatino	0.44	1.03	0.43	-1.39	0.61	-2.27*	

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

As previously reported, the percent of variance explained by child factors within the full sample totaled 17.24%. In combination, the student level predictors (e.g. age, gender, race/ethnic heritage, and eligibility for sped, disability accommodations, and ELL services) reduced the within-classroom variance by 17.24%. The empty models for Title 1 and Non-Title

1 schools were also compared with regression coefficient models containing child level predictors. Interestingly, a significantly greater proportion of the within classroom variation was explained by child factors in Title 1 schools ( $\sigma^2$  empty model -  $\sigma^2$  level 1 base model/ ( $\sigma^2$  empty model)] or (104.93-77.33/104.93 = 26.30). For Title 1 schools, 26.30% of the level 1 variability was explained by the addition of child factors included in the base model. In contrast, only 16.23% of the within classroom variability was explained by the same child factors for Non-Title 1 schools ( $\sigma^2$  empty model -  $\sigma^2$  level 1 base model/ ( $\sigma^2$  empty model)] or (93.29-78.15/93.29 = 16.23).

To investigate the explanatory contributions of the level 1 child variables, the predictors were separated into three variable groups: 1) race and ethnic factors (e.g. Asian, Black/African-American, Latino/Hispanic, Multi-Ethnic), 2) eligibility for support factors (e.g. Special education, disability accommodations, English language learning), and 3) age and gender (e.g. age at time of rating, gender). Each variable grouping was independently added to the fully unconditional model to evaluate the extent to which these variable groups explained the within classroom variability in scores. The percentage of additional variance explained by the addition of these factors to the unconditional model for Title 1 and Non-Title 1 schools are shown in Table 5.20.

**Table 5.20.** *Title and Non-Title level 1 variable group variance contributions*

	Percent of Within-Class Variance Explained	
	Title 1 Schools	Non-Title 1 Schools
Racial/ Ethnic Factors	4.71	0.54
Service Eligibility	13.89	8.22
Age and Gender	13.62	9.15

In Non-Title 1 schools, racial and ethnic factors reduce the within-classroom variance that is 83.29% of the total variability in student scores by 0.54%. Eligibility for support services explains an additional 8.22% of the variability and age and gender reduce the within classroom variance in student scores by 9.15%. A large portion of the within classroom variance remains unexplained. In Title 1 schools, the child level predictors for each variable group reduce the variance in teacher ratings of students' social and emotional competency by a greater amount than the Non-Title 1 schools. Race and ethnic factors reduce the unexplained within classroom variation in student scores by 4.71%. Support service eligibility factors explain 13.89% of the total variation in student scores at the within classroom level and age and gender explain 13.62% of the within classroom variability at Title 1 schools.

To test the statistical differences in the significance of level 1 predictor effects observed between Title 1 and Non-Title 1 schools, a level 3 variable was created to distinguish school type based on Title 1 school status. Title 1 schools were dummy coded as 1 and Non-Title 1 schools were coded as zero. Another HLM model was constructed to include cross level interactions between school sector (e.g. Title 1 or Non-Title 1) and level 1 predictors (e.g. age at rating, gender, Sped status, disability accommodation status, English language learner status, and race/ethnic heritage predictors). The results are shown in Table 5.21.

**Table 5.21.** *Effects of School Type on Child Level Predictors of SET Total*

<i>Fixed Effects</i>	Social and Emotional Total			
	<i>Coef.</i>	<i>se</i>	<i>t-ratio</i>	<i>p-value</i>
SET Total				
Intercept	52.35	0.50	105.65	0.00
Title 1	-3.98	1.33	-3.00	0.01
AgeatRating				
Intercept	0.14	0.04	3.84	0.00
Title 1 School	-0.02	0.10	0.24	0.81
ISFemale				
Intercept	5.33	0.31	17.23	0.00
Title 1 School	0.84	0.82	1.03	0.31
ISSped				
Intercept	-9.20	0.67	-13.79	0.00
Title 1 School	1.48	1.52	0.97	0.33
ISDis				
Intercept	-5.95	1.00	-5.96	0.00
Title 1 School	3.01	2.07	1.46	0.15
ISELL				
Intercept	-3.19	0.62	-5.13	0.00
Title 1 School	-1.19	1.11	-1.07	0.29
ISAsian				
Intercept	0.37	0.37	0.99	0.32
Title 1 School	3.42	1.20	2.86	0.00
ISBlack				
Intercept	-4.38	1.25	-3.51	0.00
Title 1 School	0.87	1.96	0.44	0.66
ISMultiEthnic				
Intercept	-0.50	0.58	-0.86	0.39
Title 1 School	-0.29	1.78	-0.16	0.87
ISLatino				
Intercept	-1.39	0.61	-2.27	0.04
Title 1 School	1.78	1.21	1.47	1.61

Teacher ratings yielded statistically significant differences in social and emotional competency scores between students in Title 1 and Non-Title 1 schools. On average, children at non-Title 1 schools were rated 3.98 t-score points higher than peers enrolled in Title 1 schools. School context statically significant moderated the effects of age at rating, gender, special education status, disability accommodation status, and English language learning status on teacher ratings of students' social and emotional competency. A student's race or ethnic

heritage also had differential effects based on the school context. School sector differences in teacher ratings, as previously discussed, were observed for Black/African-American, Latino, and Asian students.

At the classroom level, class size and the percent of non-White non-Asian students in a classroom were entered as grand mean centered random variables in a means-as-outcomes model (Raudenbush, 2002) in which level 1 independent variables were removed. The predicted social and emotional competency score for students in classrooms with an average number of students and with an average percentage of students belonging to non-White non-Asian racial and ethnic groups is a SET T-score of 53.41 (M = 53.41) (see Table 5.22).

**Table 5.22.** *HLM level 2 base model by social emotional total*

<i>Classroom Effects</i>	Social and Emotional Total		
	<i>Coefficient</i>	<i>se</i>	<i>t-ratio</i>
Intercept	53.41	0.46	116.70**
ClassSize	0.06	0.16	0.34
P Disadvantaged	-0.06	0.03	-2.08

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

Class size did not significantly predict teacher ratings of students' social and emotional competency. The percent of non-White non-Asian students within a classroom approached statistical significance on the SET outcome measure. For every percentage increase in the proportion of non-Asian and non-White students in a classroom relative to the grand average, a student's predicted outcome score decreased an average of 0.06 points when holding all other factors constant ( $t = -2.91$ ,  $p = 0.053$ ).

After accounting for the effects of level 2 predictors, the differences in student scores attributable to between-school differences were no longer statistically significant. The final estimation of level-3 variance components yielded a  $\chi^2$  value of 14.94 with a  $p > 0.50$ . This

value was not statistically significant, therefore, after controlling for the percent of non-White non-Asian students within a classroom and class size, variation between-schools no longer remained as a significant source of variability in the outcome measure.

Classroom level predictors were also evaluated for the Title 1 and Non-Title 1 school sub-samples. As in the full sample, level 2 covariates did not predict teacher ratings of students social and emotional competence within either of the segmented school sub-samples evaluated (See Table 5.23.).

**Table 5.23.** *HLM level 2 base model by social emotional total for Title 1 and Non-Title 1*

<i>Student Effects</i>	Social and Emotional Total						
	<i>Coef.</i>	Title 1 Schools			Non-Title 1 Schools		
		<i>se</i>	<i>t-ratio</i>		<i>Coef.</i>	<i>se</i>	<i>t-ratio</i>
Intercept	48.95	0.87	55.98	54.00	0.46	118.24	
ClassSize	0.08	0.46	0.18	0.01	0.18	0.06	
P_Disadvantaged	0.07	0.17	0.40	-0.02	0.04	-0.65	

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

At the school level, the average years of teaching experience, the percentage of students receiving free or reduced lunch, and school enrollment were entered as grand mean centered variables. The base model regression coefficients for level 3 covariates are shown in Table 5.24.

**Table 5.24.** *HLM level 3 base model by social emotional total*

<i>School Effects</i>	Social and Emotional Total		
	<i>Coefficient</i>	<i>se</i>	<i>t-ratio</i>
Intercept	52.82	0.35	152.15
Enrollment	0.00	0.00	0.45
AvgYrTeach	0.19	0.25	0.75
P_FreeRedLunch	-0.11	0.03	-3.90**

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

The percentage of students receiving free/ reduced meal services was a significant predictor of teachers' ratings of SEC. For every percentage point increase in the number of students eligible for free and reduced lunch within a school, there is an expected 0.11-point

decrease in t-score predicted in a student's score. The average years of teaching experience among teachers within a school and the number of students within a school were not significant predictors of teacher ratings of student social and emotional competency. Level 3 covariates were not investigated for Title 1 and Non-Title 1 school sub-samples because the low number of level 3 units restricted the analysis.

*Intercepts and slopes as outcome models.* In the third wave of analysis, intercepts and slopes as outcomes regression models (Raudenbush, 2002) were constructed to evaluate the contributions of predictors at the child, classroom, and school level to the variability in teacher ratings of student social emotional competency while also accounting for variability within and between classrooms, and between-schools. Using the full sample, level 2 and level 3 predictors were sequentially added to level 1 predictors. In Model 1 (shown in Table 5.20.), class size and percent of underserved students were entered as level 2 grand mean centered random covariates. After controlling for level 1 and level 2 variables, differences between-schools no longer accounted for a significant portion of variability in student scores. In Model 2, school enrollment, the average years teaching experience among school teaching staff, and the percentage of students eligible for free or reduced lunch were added to Model 1 as grand mean centered level 3 predictors. Intercepts and regression coefficients from each of these models are shown in Table 5.25. In a model comparison test, of Model 1 and Model 2, Model 2 is not a better fit to the data than Model 1 ( $\chi^2 = 3.75, df = 3, p = 0.29$ ).

**Table 5.25.** *Coefficients from HLM analyses of SET total score*

<i>Variable</i>	<i>Model 1</i>		<i>Model 2</i>	
	<i>Coefficient</i>	<i>se</i>	<i>Coefficient</i>	<i>se</i>
Intercept	53.41	0.40	53.08	0.43
Student level				
AgeatRating	0.14**	0.03	0.14**	0.04
ISFemale	5.46**	0.47	5.46**	0.35
ISSped	-8.91**	0.91	-8.90**	0.60
ISDis	-5.36**	0.89	-5.35**	0.87
ISELL	-3.58**	.052	-3.53**	0.51
ISGifted			-0.12	0.81
ISAsian	0.71	0.37	0.70*	0.31
ISBlack/African	-4.49**	0.81	-4.46**	0.94
ISLatino	-1.21*	0.54	-1.19*	0.51
ISMultiEthnic	-0.50	0.55	-0.52	0.55
Class level				
Class Size	0.05	0.13	0.01	0.14
P_Disadvantaged	-0.02	0.04	0.02	0.61
School level				
Enrollment			-0.00	0.00
AvgYrTeach			0.05	0.27
P_FreeRedLunch			-0.08	0.04

\*\*  $p < .001$ , \*  $p < .05$

In sum, after controlling for classroom and school membership, only child level variables were significant predictors of teacher ratings of students' social and emotional competence. While a sizable portion of variance is located at the between-classroom, the variables tested fail to explain the variability. The between-classroom differences remain unexplained. Even in the unconditional baseline model, the amount of variance explained by between-school differences was relatively small. The statistical significance of between-school differences was eliminated after introducing child and classroom predictors.

## Qualitative Findings

The qualitative analyses of the 15 semi-structured teacher interviews investigated teachers' thoughts pertaining to the assessment of students' social and emotional competency. The analysis explored broad patterns in teachers' response styles and investigated the ways in which teachers' experience, thoughts, beliefs, and biases influenced their assessment practices. Grounded in Bronfenbrenner's bioecological theory (Bronfenbrenner, 1995; Bronfenbrenner & Morris, 1998) and social learning theory (Baer & Bandura, 1963), the analysis also evaluated teacher references to classroom and school context characteristics that were evaluated in the quantitative analysis. The investigation examined the ways in which these influences contributed to teachers' thoughts about child, classroom, and school factors to better understand how these factors contribute to ratings of student behavior.

The four cognitive processes outlined in the cognitive response theory (Tourangeau, 2003) guided a second phase of the analysis. Interviews were analyzed to explain how teachers comprehended the measure, recalled pertinent information, formed judgements about students, and selected responses from the available response choices. The entirety of the qualitative analysis was broadly framed by the 5 principal research questions guiding the mixed methods study.

- (a) How do teachers' interpretations and responses to social and emotional rating scale measures compare with teachers working in the same or different school contexts?
- (b) What is the relationship of within classroom variability to between-classroom variability in teacher ratings of students' social and emotional competencies?
- (c) How do teachers expressed thoughts when completing a social and emotional rating scale reflect child, informant, and contextual factors in the process of rating students' social and

emotional skills and competencies? How do these factors influence a teachers' ratings of individual children?

- (d) To what extent are child, informant, and contextual factors associated with teachers' ratings of students' social and emotional competence?
- (e) How do child, informant, and contextual factors and instrument-related factors play into teachers' thinking when rating students' social and emotional competencies?

**Representativeness of participants and classrooms to full sample.** The 15 participants' average years of teaching experience ( $M = 11.13$ ,  $\text{min} = 1$ ,  $\text{max} = 29$ ) was comparable to the average years of teaching for classroom teachers in the participating district ( $M = 10.50$ ) ([www.k13.wa.us](http://www.k13.wa.us)). The percentages of teachers from racial and ethnic groups also approximated district averages. Among participants, Asian teachers comprised 13% of the sub-sample. In comparison, Asians make up 7.8% of the teacher population within the district. Latino teachers made up 6% of all teachers interviewed. Within the participating district, only 4.3% of teachers identified as Latino. White teachers were the largest racial group represented among both the interviewed teachers (80%) and teachers working in the participating district (86.4%).

**Classroom composition.** The student compositions of participating classrooms were also reasonable approximations of the full sample regarding class size, gender distributions, racial/ethnic heritage, and the overall percentages of students eligible for support services (e.g. Sped, disability accommodations, ELL services). The largest difference in the mean percentages of students represented within full sample and participating teacher classrooms was observed in the percentage of students eligible for English language learning services. Participating teachers' classrooms on average had 6.51% more students receiving English language services than

teachers' classrooms across the full sample. See table 5.26 for a visual summary of classroom composition for participating teachers and those of the full sample.

**Table 5.26.** *Comparison of participating classrooms to full sample classrooms*

<i>Variable</i>	<i>N</i>	<i>Mean</i>	<i>Min</i>	<i>Max</i>
		<u>Full Sample</u>		
Class Size	173	23.15	10.00	30.00
FP_ISFemale	173	48.93	31.60	68.20
P_ISSPED	173	6.20	0.00	25.00
P_ISDis	173	2.84	0.00	19.00
P_ISELL	173	10.89	0.00	61.90
P_ISAsian	173	37.72	0.00	89.30
P_ISBlack	173	2.97	0.00	21.40
P_MultiEthnic	173	8.53	0.00	32.00
P_Latino	173	12.83	0.00	85.70
		<u>Participating Teacher Classrooms</u>		
Class Size	15	22.93	20.00	25.00
FP_ISFemale	15	51.41	35.00	68.20
P_ISSPED	15	4.25	0.00	8.30
P_ISDis	15	3.75	0.00	4.20
P_ISELL	15	17.40	0.00	20.80
P_ISAsian	15	33.20	9.10	76.90
P_ISBlack	15	6.79	0.00	23.80
P_MultiEthnic	15	7.41	0.00	23.80
P_Latino	15	15.76	0.00	38.10

*Composition of Non-Title 1 and Title 1 classrooms compared to full sample.* Among the Non-Title 1 participating teachers' classrooms, there were on average 13.64% fewer Latinos than in Non-Title 1 classrooms in the full sample. This difference is partially attributable to the low number of Latino students in gifted classrooms. On average, the percentage of Latino students in gifted classrooms within the participating classrooms was 1.33%. The full sample of Title 1 classrooms on average included 7.99% more Asian students than the classrooms of participating teachers. Comparisons of classroom compositions for participating teachers' classrooms and the

classrooms included in the full sample for Title 1 schools and Non-Title 1 schools are shown in Table 5.27 and 5.28, respectively.

**Table 5.27.** Comparison of Non-Title 1 participating classrooms to full sample classrooms

<i>Variable</i>	<i>N</i>	<i>Mean</i>	<i>Min</i>	<i>Max</i>
		<u>Full Sample</u>		
Class Size	146	23.48	10.00	30.00
FP_ISFemale	146	49.11	31.60	68.20
P_ISSPED	146	5.83	0.00	25.00
P_ISDis	146	2.51	0.00	16.00
P_ISELL	146	7.09	0.00	21.70
P_ISAsian	146	40.73	0.00	89.30
P_ISBlack	146	1.94	0.00	19.20
P_MultiEthnic	146	8.93	0.00	32.00
P_Latino	146	19.24	0.00	36.00
		<u>Participating Teacher Classrooms</u>		
Class Size	9	24.33	20.00	26.00
FP_ISFemale	9	52.94	35.00	68.20
P_ISSPED	9	1.70	0.00	8.30
P_ISDis	9	0.91	0.00	4.20
P_ISELL	9	9.78	0.00	20.80
P_ISAsian	9	45.86	17.40	76.90
P_ISBlack	9	2.37	0.00	9.10
P_MultiEthnic	9	7.06	0.00	17.40
P_Latino	9	5.60	0.00	13.60

**Table 5.28.** Comparison of Title 1 participating classrooms to full sample classrooms

<i>Variable</i>	<i>N</i>	<i>Mean</i>	<i>Min</i>	<i>Max</i>
		<u>Full Sample</u>		
Class Size	27	21.37	14.00	25.00
P_ISFemale	27	47.94	36.40	59.10
P_ISSPED	27	8.19	0.00	18.20
P_ISDis	27	4.64	0.00	19.00
P_ISELL	27	31.44	14.30	61.90
P_ISAsian	27	21.46	0.00	54.50
P_ISBlack	27	8.53	0.00	21.40
P_MultiEthnic	27	6.34	0.00	14.30
P_Latino	27	37.13	22.70	85.70
		<u>Participating Teacher Classrooms</u>		
Class Size	6	21.17	20.00	22.00
FP_ISFemale	6	49.10	38.10	59.10
P_ISSPED	6	7.95	04.80	14.30
P_ISDis	6	8.02	0.00	23.80
P_ISELL	6	28.53	22.70	36.40
P_ISAsian	6	13.47	9.10	19.00
P_ISBlack	6	11.03	9.50	13.60
P_MultiEthnic	6	6.35	0.00	14.30
P_Latino	6	31.00	27.30	38.10

Descriptors of student characteristics within each of the participating teachers' classrooms are shown in Appendix C.

**Classroom ratings.** The classroom DESSA-Mini SET mean scores for the 15 teacher interview participants reasonably approximated those of the full sample. Participant teachers' DESSA-Mini SET classroom means ranged from 42.50-59.20 with an overall sub-sample mean score of 52.93 (M = 52.93). The grand classroom mean for the full sample differed by less than a single t-score point (M = 53.31). Differences were observed in the range of classroom mean scores between participant teachers and the full sample. Classroom mean scores for participating teachers and the full sample of teachers shared a minimum classroom mean score

of 42.50 (minimum classroom mean = 42.50); however, the maximum classroom mean score observed among the full sample (M = 68.00) exceeded the maximum classroom mean score of the participating teachers (M = 59.20). See Table 5.29 for a summary of classroom mean scores and demographic information for participating teachers. In addition to classroom means, teacher years of teaching experience, gender, race are provided. Class size, the range of responses utilized by each participant, the within-student mean variance, and the percent of students rated the same across all items also provided to describe teachers' responses to the rating scale measure. Teachers working in Title 1 and Non-Title 1 schools are tabled separately to facilitate comparison of teacher behaviors among educators working in different contexts.

**Table 5.29.** *Teacher demographics and descriptive statistics for classroom ratings*

<i>Int.</i>	<i>Years</i>	<i>Gender</i>	<i>Race</i>	<i>N</i>	<i>Mean (sd)</i>	<i>Range</i>	<i>M_Var</i>	<i>0_Var</i>
					<u>Title</u>			
2	1	F	White	21	59.29 (11.75)	1-4	0.38	23.81
3	13	F	White	21	56.19 (10.55)	1-4	0.35	14.29
4	5	F	White	22	57.64 (11.95)	0-4	0.42	13.64
9	10	F	Latina	22	52.91 (11.46)	0-4	0.74	0.00
12	9	M	White	20	42.50 (06.72)	0-3	0.26	10.00
14	14	F	White	21	53.52 (12.85)	1-4	0.36	19.05
					<u>Non-Title</u>			
1*	15	F	White	25	52.92 (07.47)	1-4	0.19	16.00
5	13	F	White	25	54.72 (12.06)	0-4	0.41	24.00
6	18	M	White	26	56.31 (12.87)	0-4	0.31	30.77
7*	2	F	White	20	43.70 (03.54)	1-3	0.39	05.00
8*	5	F	White	26	48.85 (04.99)	1-3	0.15	34.62
10	29	F	White	22	55.36 (13.18)	0-4	0.38	18.18
11	25	F	White	24	54.33 (14.36)	0-4	0.18	41.67
13	3	F	Asian	24	51.54 (11.28)	1-4	0.18	33.33
15	5	F	Asian	25	54.12 (08.12)	0-4	0.55	0.00

\* self-contained gifted classroom, M\_Var = the average variation contained within-student ratings, 0\_Var = % of students in class that had no variability in ratings across items

*Classroom ratings of Non-Title 1 interview participants.* The Non-Title 1 participating teachers' DESSA-Mini classroom mean t-scores ranged from 43.70-56.31 (min = 43.70, max = 56.31) with a mean of 54.44 (M = 54.44). The overall classroom mean among Non-Title 1 teacher participants was comparable to the grand classroom mean observed among Non-Title 1 teachers in the full sample (M = 53.97). The range in classroom means among participating Non-Title 1 classroom teachers was narrower (min. = 43.70, max. = 56.31) than the range among Non-Title 1 classrooms in the full sample (min. = 43.70, max. = 68.00) and was limited at the upper end of the range by 8.8 points.

*Classroom ratings of Title 1 interview participants.* The participating Title 1 teachers' classroom means matched range observed among Title 1 teachers in the full sample (min = 42.50, max = 59.20). The participating teachers working in Title 1 schools on average rated their students more favorably than Title 1 teachers in the full sample and the overall mean classroom mean score among participating Title 1 teachers was 3.91 points higher (M = 53.66) than the overall classroom mean among Title 1 teachers observed in the full sample (M = 49.75). See Tables 5.30 for a summary comparing participating teachers to the full sample for Title 1 schools, Non-Title 1 schools, and the total.

**Table 5.30** *Comparison of participating teachers to full sample*

	<i>Social and Emotional Total</i>					
	<u>Participating Teachers</u>			<u>Full Sample</u>		
	N	Mean	Range	N	Mean	Range
Title 1	6	53.66	42.50-59.20	27	49.75	42.50-59.20
Non-Title	9	52.44	43.70-56.31	146	53.97	43.70-68.00
Total	15	52.93	42.50-59.20	173	53.31	42.50-68.00

**Students selected for teacher ratings.** Each of the teacher participants rated three students. The characteristics for students selected as having strong, typical, and need range social and emotional skills are shown in Tables 5.31, 5.32, and 5.33, respectively.

**Table 5.31.** *Child characteristics of students rated in strength range*

<i>Characteristic</i>	<i>Title 1</i>	<i>Non-Title</i>	<i>Total</i>
Gender	6 F, 0M	7 F, 2 M	13 F, 2 M
Race/Ethnic Heritage			
Asian	1	3	4
Black/AA	--	--	--
Latino	--	--	--
MultiEthnic	3	1	4
White	2	5	7
Services			
ELL	1	1	2
Sped	--	--	--
Dis	--	--	--
Gifted	--	3	3
None	4	7	11

**Table 5.32.** *Child characteristics of students rated in typical range*

<i>Characteristic</i>	<i>Title 1</i>	<i>Non-Title</i>	<i>Total</i>
Gender	4 F, 2 M	4 F, 5 M	8 F, 7 M
Race/Ethnic Heritage			
Asian	--	5	5
Black/AA	1	--	1
Latino	3	--	3
MultiEthnic	--	2	2
White	2	2	4
Services			
ELL	2	1	3
Sped	--	--	1
Dis	--	--	--
Gifted	--	3	3
None	4	5	9

**Table 5.33.** *Child characteristics of students rated in need range*

---

<i>Characteristic</i>	<i>Title 1</i>	<i>Non-Title</i>	<i>Total</i>
Gender	1 F, 5 M	5 F, 4 M	6 F, 9 M
Race/Ethnic Heritage			
Asian	1	3	4
Black/AA	--	--	--
Latino	2	1	3
MultiEthnic	--	3	3
White	3	2	5
Services			
ELL	3	1	4
Sped	1	--	1
Dis	1	--	1
Gifted	--	3	3
None	1	6	6

---

**Gender.** Among students selected for the strength range, females are overrepresented and comprise 86.67% of the children selected as exhibiting strong social and emotional skills. Females are slightly underrepresented among students within the need range and make up 40% of students selected as exhibiting deficient social and emotional skills. This pattern is consistent with the quantitative findings from the current study suggesting that student gender is associated with significant differences in teachers' ratings of students' social and emotional competency. On average, teachers rate females more favorably than males.

**Race and ethnic heritage.** Teacher selections of students for the strength, typical and need ranges is consistent with quantitative findings in which being Black/African-American or Latino/Hispanic significantly predicted lower ratings on the teacher measure of social and emotional competency. Black/African-American and Latino students were underrepresented among children selected for the strength range. These racial and ethnic groups represented 5.83% and 15.76% of all students in the participating teachers' classrooms, respectively, yet none of the

15 students selected for the strength range were Latino or African-American. Within the typical and need ranges, the racial and ethnic heritage of selected students more consistently tracked the proportions of students from these groups within the participating classrooms. The overall proportion of Asian students identified within the strength, typical, and need range was stable across the three social and emotional competency categories, with many of the Asian students being selected by Non-Title 1 teachers. The higher than expected ratio of Asians within the need range may be explained by the high percentages of Asian students in many of the Non-Title 1 classrooms. Four of the 9 Non-Title 1 classrooms are comprised of over 50% Asian students.

***Support services.*** Only 2 of the students selected as representative of the strength range received any type of support services. These students were eligible for English language learner services. The percentage of students receiving ELL services was inversely associated with the higher levels of social and emotional competency, with double the number of ELL students represented in the need range ( $n = 4$ ) than in the strength range ( $n = 2$ ). The number of students receiving gifted services remained constant as these students were enrolled in gifted classrooms in which all students received curricula designed for intellectually talented students. Only 3 students selected for teacher ratings received services for diagnosed disabilities. Two students received special education services and one student received disability accommodations.

***Response patterns in teachers' ratings.*** Three cognitive response style prototypes emerged from teachers' ratings of student behavior. Teachers responses differed in the length of response, the ways in which they recalled information about students, and the level of detail used to describe individual students, or their behaviors. The three response style prototypes include: 1) responses reliant on impulse and general impressions, 2) responses reliant on reflective thought and summative experience, and 3) responses reliant on reflective thought and specific

experiences (see Fang, 1996 for discussion of teacher judgments). Descriptions of these prototypes and how they are associated with teacher characteristics and school contexts are reported below.

***Responses reliant on impulse and general impressions.*** Informants of this prototype provided quick responses of limited length that offered global generalizations of a student, personality attributes, character traits, or no description at all. For example, teachers described students using phrases like, “she’s just a normal great kid” (Interview #10), “she is clearly above grade level” (Interview #2). Others provided rating descriptor categories or numbers, without explanation to substantiate the rating. One teacher, for example, responded “Contributing to the group. She’s a 3” (Interview #6), and another shared, “She’s *always* paying attention” (Interview #5).

The impulse and general impressions respondents' ratings of students provided limited--if any--reference to specific experiences, or observations of students to support their response selections. Generally, response selections were made quickly and stated with certainty. Of the 27 quotations in teacher interviews coded for deliberating a response, only a single instance appeared in interviews with the 4 teachers exhibiting the impulse and general impressions response style.

In the current sample of teachers, the years of teaching experience, the school environment, and/or class size was associated with teacher response styles. Three quarters of teachers exhibiting the impulse and general impressions response style had 10 or more years of teaching experience ( $M = 15.25$ ) and were teaching in Non-Title 1 schools. The average class size among these teachers was 23.5 ( $M = 23.5$ ) students. Within the full sample of participating teachers, these teachers were some of the most experienced. Each of these highly-experienced

educators referenced the reasonableness of relying on teacher opinion when evaluating students' social and emotional skills. One teacher recommended selecting rating responses based on "general teacher opinion" (Interview #10), rather than distinguishing differences among frequency ratings. Another participant more explicitly stated that it was "better to leave it to the teacher to judge [the frequency of student behaviors] because you just know" (Interview #5).

The fourth teacher, appearing to exhibit an impulse and general impressions response style, was a first-year educator working in a Title-1 school. The length, content, and nature of her responses may have been influenced by a relative lack of teaching experience, rather than a reliance on personal instinct and intuition that others exhibiting this response style had developed over years of teaching. In contrast to the certainty regarding the appropriateness of teachers evaluating students' social and emotional skills expressed by the more experienced teachers, the novice teacher expressed concern about her ability to accurately evaluate student behavior.

"Well how can we, how can we evaluate them? Students? Children? When they are placed in a setting, like a classroom, with me and I am not a psychologist trained to do that (Interview #2).

Examples of impulse and general impressions response style ratings for individual student behaviors are provided below:

- a) Does she do something nice for somebody? All the time. *She's just kind and quiet and thoughtful. She is someone who is very positive* (Interview #5).
- b) Contributes to the group efforts. Yes. *She loves to work with others. Loves to be part of a team* (Interview #10).
- c) And pay attention, not at all. He has an IEP. *He has attention deficit, but we know there's a lot of other stuff going on there* (Interview #6).

***Responses reliant on reflective thought and summative experience.*** Informants reliant on the reflective thought and summative experiences provided brief general accounts of students

that typically lacked specific examples of observed student behavior. Descriptions were occasionally nuanced and described variation in performance based on academic context or setting demands (e.g. large group vs. small group). In selecting rating responses, respondents occasionally articulated consideration of multiple options and deliberated response selections.

Examples:

- a) Do something nice for somebody, yeah, I don't think he goes out of his way. He's not a mean kid, but I don't know if he drops little notes for people or would notice if somebody needed something and do it for them. Occasionally, he will certainly share things and do an extra part in group work, so I would say occasionally (Interview #8).
- b) Contributes to group efforts. I would give this child a 3 in contribute to group effort because she's a very flexible person and she works well with other teammates. I wouldn't say that she has strong leadership skills in terms of taking charge when she's in a group, but is very flexible and works well with no matter other person she's grouped with (Interview #13).
- c) Pay attention. She's pretty much always on task I would say depending on the content area and how much talking is required... frequently to very frequently I guess (Interview #12).

The years in teaching experience among the 6 teachers exhibiting this response style ranged from 2 to 25 years ( $M=10.00$ ); however, 3 of the 5 teachers had taught for 5 or fewer years. These educators worked in mid to high income school contexts. Five of the 6 teachers taught in the same school and half of the group taught in classrooms for gifted students. The class sizes among this group of teachers was comparable to those of the intuition and impulse group, with an average of 23.8 students in each class ( $M = 23.8$ ).

***Responses reliant on reflective thought and specific experiences.*** Informants of this response style provided detailed descriptions of student behavior, often citing specific incidents or examples. Their descriptions were often nuanced and described variation in behavior dependent upon academic contexts or setting demands. Many of the ratings also included

detailed information about students' behavioral tendencies, reflecting substantial knowledge of student capabilities. When selecting a response from those provided, the respondent often considered multiple options and sometimes deliberated the response options provided.

Examples:

- a) Do something nice for somebody. I would say occasionally. He does seem to have a little ... I think there's more in him, because I do see glimpses of it very deliberate. *'You did a great job, today!'* If somebody is struggling a little bit, or maybe they don't usually get affirmation, occasionally he will be that person that says, *"that was a great answer."* Or *"great work."* Overall, I do feel he's still kind of in that stage of, 'this is just my world, right now.' More thinking about himself. If I were to sit back and say, "let's think about somebody in this classroom that needs us to do something nice for them," he would absolutely be the one that would say, "okay, this person needs me to do this." Not on his own, yet (Interview #9).
  
- b) Contribute to group efforts? I would put her as a 3 on this one as well. Kind of for the same reasons about do something nice for somebody else. Definitely *does what she is supposed to do in the group*, but don't see her being assertive in trying to really get the group on track. With that, at times, like if she has a really outspoken group member, she often sort of (what is that word) lets them take over and sometimes to the groups' detriment. *For example, there was this one time and they were working on a STEM project and she knew what she was supposed to do. I know because I asked her and um one of her more outspoken group members really got the group off doing the wrong work and not doing what they were supposed to be doing. And their group essentially failed on the task. And I said to this student, you knew exactly what to do and she said I know I just didn't know what to say. Didn't feel confident enough to say "We should do this."* When she did know and then was very upset when her group was not successful. So, she does contribute you know, but she is not assertive I guess. So, what did I say, a three for her (Interview #4)?
  
- c) The attention? I would say ... *It's hard, because different subjects she pays more attention in.* I would give her a 3. *She's very active, so sitting still for a long time is difficult.* If my expectation is, and I don't really do this, but if I were to say, "I need to see your eyes. I need to see this, this, and this" she's very responsive to that. If left to her own devices, she pays attention to things she's interested in. *If something's going on too long, or she's not interested in it, obviously, that's a little more difficult.* Overall, she frequently pays attention to both myself and if somebody's talking in a group. She's also very attentive to the social group conversations (Interview #14).

Eighty-three percent of teachers exhibiting the reflective thought and specific evidence response style taught in Title 1 schools. The teaching experience within the group ranged from 5 to 14 years ( $M = 9.33$ ). The class sizes among teachers within this group on average contained two fewer students ( $M = 21.8$ ).

**Response style and ratings--*Range*.** Differences in teachers' response styles were associated with assessment practices when rating students' social and emotional competency. Teachers reliant on impulse and general impressions and on reflective thought and specific experiences were more likely to utilize the full range of response options when rating student behaviors, than those who relied on reflective thought and summative experiences. In contrast, 80% of the teachers whose responses appeared reliant on reflective thought and summative experiences utilized a restricted range of response options, with two teachers only using only mid-range responses including 'rarely,' 'occasionally,' and 'frequently' response selections. Two other teachers within this group utilized 'rarely' as the lower limit of her response selections. The teachers' range of responses per response type are shown in Table 5.34.

**Table 5.34.** *Teacher demographics and statistics for classroom ratings by response type*

<i>Int.</i>	<i>Years</i>	<i>Gender</i>	<i>Race</i>	<i>N</i>	<i>Mean (sd)</i>	<i>Range</i>	<i>M_Var</i>	<i>0_Var</i>	<i>Title</i>
<u><i>Impulse and general impressions</i></u>									
2	1	F	White	21	59.29 (11.75)	1-4	0.38	23.81	Y
5	13	F	White	25	54.72 (12.06)	0-4	0.41	24.00	N
6	18	M	White	26	56.31 (12.87)	0-4	0.31	30.77	N
10	29	F	White	22	55.36 (13.18)	0-4	0.38	18.18	N
<u><i>Reflective Thought and Summative Experience</i></u>									
1*	15	F	White	25	52.92 (07.47)	1-4	0.19	16.00	N
7*	2	F	White	20	43.70 (03.54)	1-3	0.39	05.00	N
8*	5	F	White	26	48.85 (04.99)	1-3	0.15	34.62	N
11	25	F	White	24	54.33 (14.36)	0-4	0.18	41.67	N
13	3	F	Asian	24	51.54 (11.28)	1-4	0.18	33.33	N
<u><i>Reflective Thought and Specific Experience</i></u>									
3	13	F	White	21	56.19 (10.55)	1-4	0.35	14.29	Y
4	5	F	White	22	57.64 (11.95)	0-4	0.42	13.64	Y
9	10	F	Latina	22	52.91 (11.46)	0-4	0.74	0.00	Y
12	9	M	White	20	42.50 (06.72)	0-3	0.26	10.00	Y
14	14	F	White	21	53.52 (12.85)	1-4	0.36	19.05	Y
15	5	F	Asian	25	54.12 (08.12)	0-4	0.55	0.00	N

\* self-contained gifted classroom, M\_Var = the average variation contained within-student ratings, 0\_Var = % of students in class that had no variability in ratings across items, Title = Title 1 school (Y = yes, N = no)

***Within-student variability.*** The within-student variation in teacher ratings was also associated with teacher response patterns. The within-student mean variance across the three response prototypes ranged from 0.22 to 0.45. Teachers reliant on the reflective thought and summative experiences were more likely to rate a student similarly across each of the 8 items on the DESSA-Mini than those who relied on impulse and general impressions or specific experiences. The mean within-student variance in teacher ratings among participants reliant on summative experiences was 0.22. In comparison, the student ratings among teachers reliant on impulse and general impressions and reflective thought and specific experiences yielded mean within-student variances of 0.37 and 0.45, respectively. One plausible explanation for the lower

within-student variation observed among respondents reliant on reflective thought and summative experience, is the tendency of these raters to utilize a reduced range of response options. Eighty percent of teachers reliant on reflective thought and summative experience rated students using a narrower range of responses than provided on the measure.

The mean percentage of student ratings in which there was zero within-student variability varied widely across the three response prototypes. Teachers reliant on impulse and general impressions completed ratings in which there was no variability in 24.19% of their student ratings, with no reported differences in behavior across the 8 behaviors assessed. Among teachers reliant on reflective thought and summative experience, an even larger percentage of students (27.53%) were rated with the same response choice across all behaviors assessed (e.g. ‘frequently’). In contrast, teachers who responded using reflective thought and specific experiences generated ratings in which only 9.50% of student ratings contained no variability across items. In sum, teachers who recall specific incidents and examples of student behavior when evaluating students’ social and emotional competencies utilized a wider range of response choices within individual student ratings, possibly signaling a more accurate measure of observed student behaviors. In contrast, those relying on impulse and general impressions and a tendency to focus on global personality traits or characteristics may be less precise in their evaluation of specific behaviors.

**Teacher cognitive response styles and perceived comprehension.** Teachers’ cognitive response patterns were unrelated to their perceived comprehension of behavioral terms and phrases used on the rating scale measure. Teachers’ self-reported level of comprehension for each of the behaviors probed (e.g. paying attention, contributing to group efforts, doing something nice for somebody) was averaged these individual teacher averages were included in

the summative average for each of the response prototype groups. The overall average across groups was at the 'comprehend' category level ( $M = 3.1$ ) with the comprehension across the three groups being roughly equivalent. On the 5- point Likert-type scale (0 = no comprehension, 1 = uncertain of comprehension, 2 = somewhat comprehend, 3 = comprehend, 4 = fully comprehend), the impulse and general impressions, reflective thought and summative experience, and reflective thought and specific experience response style groups yielded average comprehension scores of 3.10, 3.10, 3.22 respectively. Thus, the response style that teachers employed when rating students appears unrelated to their perceived comprehension of the rating scale items themselves.

**Teacher cognitive response styles and confidence in rating behaviors.** Teachers' confidence in their ability to accurately rate the frequency of student behaviors was appreciably lower than their comprehension of behavioral terms. While participants claimed to understand what it means to pay attention, or do something nice for somebody, they were less confident in their ability to accurately rate how frequently students exhibit these behaviors. Participants noted challenges with a) the visibility of these behaviors among some groups of students, b) a lack of clarity in how to interpret the frequency ratings, and c) variation in the behavioral standards held by different teachers. These challenges compromised their confidence that their own ratings accurately described student competencies.

Several teachers shared that it is more difficult to rate quiet students and students with more limited language skill because they are less visible in the classroom. In responding to a question about how a teacher distinguishes between frequency ratings, one participant responded,

It's sometimes...[difficult] because I don't see as much. I don't hear as much. I don't have enough information or feedback there to necessarily break it down into one [rating category] or the other (Interview # 1).

Another teacher noted that quiet students may be less likely to utilize their skills,

Some kids are easier to think about like that than others.... Because maybe you may have more visibility in what they are doing.... like sometimes for quiet kids it's hard for them to show that they really do participate in groups productively. Whereas, other kids may be really demonstrative either positively or negatively and that's sometimes easier to observe when you are trying to work with all of your kids (Interview #4).

Many teachers expressed how difficult it is to attach frequency terms to their daily observations of students within the classroom. One teacher questioned the practical meaning of the rating categories, "when I'm looking at these ratings, I don't know what 'very frequently' would be. Is that all the time" (Interview #8)? Some participants also shared concerns that raters have varying levels of tolerance and expectations for student behavior.

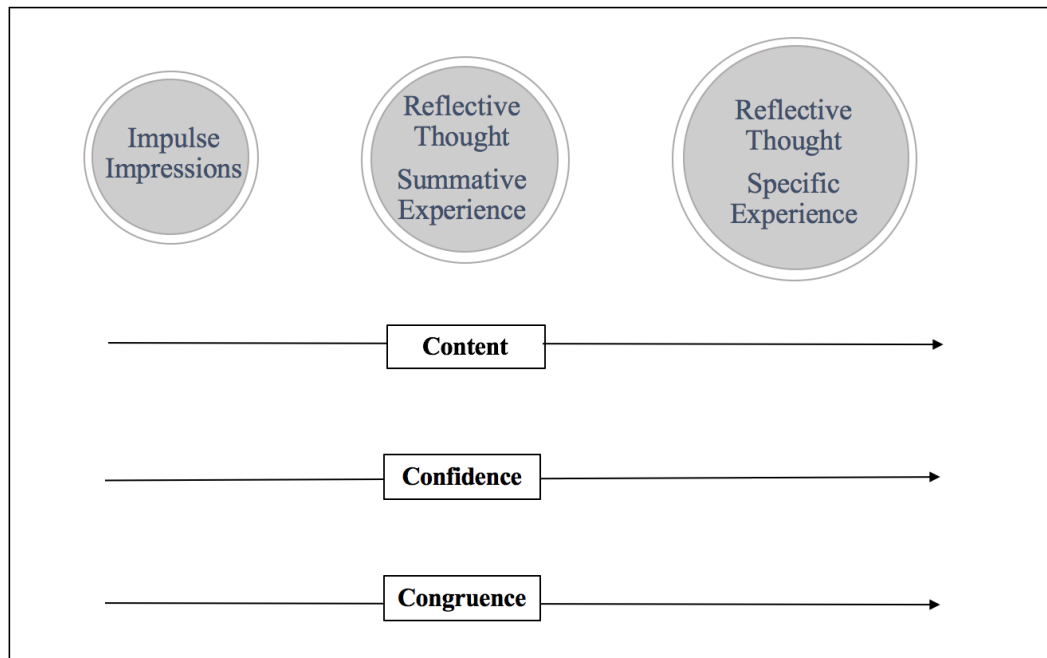
Some people will say they have a gazillion red flag kids and some people will have all green. They'll be like 'Okay so is that means that your definition of what these things mean is not what mine mean' or the tolerance of the teacher is extremely high, or extremely not high. Is it because we don't rate the questions the same or is because our tolerance is different? I don't know. That's one of my questions (Interview 10).

For the entire sample of participants, teachers' average level of confidence in their ability to accurately rate the frequency of student behavior fell between 'somewhat confident' and 'confident' ( $M = 2.48$ ). The impulse and general impressions group and the reflective thought and summative experience group averaged 2.17 and 2.40, respectively. Confidence in the ability to accurately rate the frequency of student behaviors was highest amongst teachers who exhibited a reflective thought and specific experiences cognitive response style. This group's confidence average was closer to 'confident' and averaged to 2.77 on the 0-4 Likert-type scale (e.g. 0 = not confident to 4 = fully confident). In sum, the more specific informants were in their response style, the more confident they were in the accuracy of their ratings. A more

thorough analysis of teachers' comprehension of behavioral terms and of frequency ratings will follow in the instrument response factors section of this paper.

The differences in cognitive response styles and the assessment practices associated with these styles offers some plausible explanations for the between-classroom variability observed in the current study. The quantitative analysis revealed that within the full sample, approximately 15.6 percent of the variation in student scores is located at the classroom, or teacher level. Yet, neither of the classroom predictors tested in the model (e.g. class size or proportion of disadvantaged student) significantly predicted teacher ratings of students' social and emotional competency. The qualitative analysis offers a plausible explanation for some of the between-classroom variability. The analysis revealed significant differences in the ways in which teachers rated students. The three response prototypes that emerged from teacher interviews revealed significant differences in the ways in which teachers rated students. These approaches differed in the types of information that teachers relied on and the ways in which they retrieved experiences. A distinguishing characteristic among these approaches was the extent to which the response style relied on specific observable events. Teachers whose response style was reliant on reflective thought and specific experiences provided more specific content about their students. These educators were more likely than those reliant on other response styles to provide information reflecting a deep knowledge about a student and experiences observing a behavior. Informants reliant on reflective thought and summative experience were more likely to distinguish capabilities across the behaviors rated producing more contrast within individual student ratings. Specificity was also associated with higher levels of confidence in rater's ability to accurately rate student behaviors and with the level of congruence between teacher's personal

notions of behavior and their student ratings. As specificity in the rating process increased, so did the content, confidence, and congruence in student ratings (See Figure 5.1).



**Figure 5.1.** *Cognitive response types and associated response qualities*

The next phase of analysis investigated instrument-related response factors by exploring teachers’ comprehension of the rating scale prompts, their retrieval of information relevant to the prompt, and their decision processes when selecting response options to rate student behaviors.

**Instrument-related factors--Comprehension.** Three DESSA-Mini prompts were probed to learn more about teachers’ thought processes regarding the rating scale measure (e.g. paying attention, contributing to group efforts, doing something nice for somebody). Teachers’ comprehension of each of the probed behaviors revealed that teachers hold variable notions about behaviors and, thus ‘look for’ different indicators that a behavior was enacted. Teachers’ comprehension of what it means to ‘pay attention’ is described here to illustrate differences in the ways in which teacher comprehend a standardized prompt. What follows is an analysis of teachers’ responses to the question, “What does it mean to pay attention?”

*Defining what it means to pay attention.* Elementary teachers are continuously soliciting, monitoring, and reinforcing students' attention to optimize learning. They intentionally direct students' attention toward specific focal activities (e.g. teacher providing directions) and away from other activities (e.g. off-topic conversations with their peers). Many elementary teachers recognize the active role the teacher plays to engage and maintain student attention. One teacher noted, "we do so much to make sure that kids are somehow able to pay attention" (Interview #12). As teachers engage in practices intended to increase student attention, they informally assess whether students are attending.

Many teachers struggled to define what it means to pay attention. The question was often followed by long pauses, gasps, or exclamations of struggle or difficulty. One teacher reasoned that "paying attention could mean multiple things" (Interview #1) and another explained that because paying attention "looks different for different kids" (Interview #6) it can be hard to define. In responding to what it means to pay attention, teachers identified more than 30 distinct behaviors that fell into 4 broad categories of behavior: 1) *attending to directions, conversation, and/or instruction*; 2) *applying understanding through written work or verbal participation*; 3) *displaying specific body language*; and/or 4) *attending to the emotions of self, and/or others*. These themes appeared in various combinations and with variable intensity across the 15 interviews. In one teacher interview, all 4 categories of attention behaviors were represented. Two teachers offered notions of attention that included categories 1-3, and 5 other teachers described attending behaviors represented in categories 2 and 3, leaving 8 unique combinations of behavior categories employed by the 15 participating teachers.

The category of attending to directions, conversation, and/or instruction was the most widely agreed upon theme with 13 of the 15 teachers identifying behaviors within this category.

The participants used a range of language to describe attending behaviors within this thematic category including “listening (Interview #8),” “getting directions,” (Interview #1) “attending to what is asked of you,” (Interview #3) “trying to understand”, (Interview #10) and “attending to other people speaking.” (Interview #5) Most teachers’ notions about paying attention included the processing of verbally presented information. Yet, a teacher’s assessment about whether a student is attending based on this notion, requires the observer to make inferences about students’ internal cognitive processes based on visible student behavior.

A smaller, but sizable group of teachers described paying attention as the ability to apply what was taught through written work or verbal responses. Two-thirds of participating teachers described a reliance on observable student behaviors in which students were applying or using learning as evidence of attention. Teachers described paying attention students as those who “follow through to complete the task,” (Interview #1) “answer and ask questions,” (Interview #7),” “respond when listening” (Interview #7) and “create some sort of output with information (Interview #12). These teachers implied that if students paid attention, they would acquire the skills necessary to respond verbally or apply their learning to written work.

Although this subset of teachers relied on verbal responses, or the application of learning to written work to demonstrate understanding as evidence of paying attention, others emphasized that understanding is not a necessary condition of attending. These differing perspectives are contrasted in two teachers’ definitions of what it means to pay attention. One teacher, offered

Paying attention means knowing, understanding what was explained and being able to do what’s expected of you later (Interview #11).

In contrast, another explained,

[Paying attention] ...means that a student is processing what I am saying. It *doesn’t* mean a student understood what I said, but they are trying to connect it to what they know. They’re trying to understand (Interview #10).

For the first teacher, understanding follows paying attention. If a child attends, he will understand. In contrast, the latter teacher believed that understanding is not always a by-product of attending and thus, she held a different standard for her students, and likely looks for other indicators suggestive of attending behavior.

Teachers' personal notions of paying attention reflected teachers' implicit beliefs regarding the relationship between students' attention and understanding. Teachers who believed that attention directly leads to understanding may fail to consider the language, learning, and emotional barriers students face that may compromise understanding, even while intently attending to instruction. Teachers may also overestimate the attention of students who possess content knowledge obtained prior to instruction, and those who quickly comprehend concepts and disengage from attending instruction. In each of these instances, a student may be able to demonstrate understanding; however, the skills may have been acquired without ever having attended to the current lesson.

In the present study, teachers working in Title 1 schools were less likely to require evidence of understanding—through the application of knowledge in written work or verbal responses--as evidence of attending. Of the 5 teachers, whose notions of paying attention did not include the application of knowledge, three taught in low income culturally and linguistically diverse schools. In contrast, 5 of the 6 teachers from high income schools identified behaviors in the *applying understanding through written work or verbal participation category*. The findings suggest that the school contexts influence the beliefs that teachers hold about learning and the criteria they employ when evaluating student attention.

A similar divide along school contexts was observed in teachers' beliefs about whether students' body language necessarily reflected their attention. Teachers in Non-Title 1 schools

were more likely to require physical behaviors such as eye contact, head nodding, and body posture as indicators of attending than teachers teaching in low SES schools. Of the 5 teachers identifying some form of body language in their definitions of paying attention, 4 taught at high income schools with fewer students from underserved racial and ethnic groups. These teachers held more stringent expectations for students. One teacher emphatically described what paying attention meant to her:

Pay attention, your eyes are on me. You are seated appropriately. Your body language is showing you are interested. You are awake. Your head is not on your desk. Your hand is not in the air (Interview #5).

This teacher was representative of others who held strong beliefs about the relationship between body language and attention. For this participant, and several others, students who were paying attention displayed a common set of behaviors (e.g. eye contact, body posture, head nodding) valued by the teacher and understood to reflect internal cognitive processes.

In contrast, 5 of the 6 teachers working in Title 1 schools explicitly stated that physical behaviors such as eye contact and body posture were not indicators of attention. One teacher clarified,

[Paying attention] ...means that you are engaged in the task at hand. So, it doesn't mean that you are being compliant, sitting there and looking at me because that doesn't mean you are paying attention at all (Interview #2).

Teachers working in low income school contexts held a more expansive notion of how attending behaviors and actively negated commonly held notions of behaviors that may indicate a student is paying attention. Per this more expansive perspective, paying attention did not require the student to be looking at the teacher, sitting in his seat, or, or be doing something specific with his hands (Interview #14).

There is also evidence in the teacher reports that for some teachers, their notions of what it means to pay attention have shifted over time and may stand counter to their training. One teacher explains that her thoughts about what it means to pay attention conflicted with her training as an ‘educator in our system.’ She concluded that that some of the theory that she learned in her teacher training had been challenged by her daily experience of teaching students with diverse needs and abilities.

Well, here's the thing. Again, I'm definitely tapping into that training I've received and just that I've been a part of, in that as educators in our system we are looking for the student with the pencil, with the notebook, with the book, and writing, and eyes on the teacher, but there's so much happening inside that child, you know, inside one, the rigor of what we're expecting each student to do, the rigor that I'm expected to have, because of what I'm supposed to get through for the year, a curriculum that isn't necessarily made for my students or my students in mind (Interview #9).

This teacher accounted for what is happening ‘inside’ the child and highlights the pressures placed on the child to meet school expectations. In recognition of these differences in her students’ capabilities, histories, and needs, she held a more expansive view of what attending looks like for her students.

Another teacher explicitly stated that his thinking over the course of his teaching experience had shifted because of his teaching experience. When asked what it means to ‘pay attention,’ he responded:

I’ve gotten a lot more lenient with this. I used to think paying attention was sitting hands still, eyes on. I’m noticing more and more this is, with this group especially, the energy is so high. I’ve just had to pull out all the stops. You’ve got kids fully lying on the floor at the beginning of the year. I was like ‘Are you kidding me? What are you doing?’ Then talking to them is like ‘What did you hear me say?’ They had it. I’ve got one kiddo who just is literally staring off into space, but when I ask her, especially during the read aloud, she has got everything. It really has shifted what I thought about what it means to pay attention (Interview #12).

The experiences of teaching ‘high energy’ students challenged this teacher’s notions of what it meant to pay attention. His revised perspective allowed for a wider range of behaviors that may

co-exist with a child attending to instruction (e.g. doodling, staring into space, moving). The contrast in teachers' notions of what paying attention looks like is dramatic. In one classroom, the teacher expected students to be sitting with a certain posture with their eyes on the speaker, while in another classroom in the same district, the instructor allowed for the possibility that a child rolling on the floor, or staring off into space may also be attending to instruction. The contrast in these thoughts and beliefs about student behavior were influenced by the teachers' experiences. Teachers who may have had similar pre-service teacher training developed different beliefs based on the realities and constraints of the environments in which they taught and the students' needs and capabilities within those contexts.

Teachers also interpreted the two other behavioral terms probed in distinct and discrepant ways. When asked what it means to 'contribute to group efforts,' teachers provided numerous unique definitions that were aggregated into 8 categorical groups. Two categories focused on the products of group contributions: 1) advancing a goal, and 2) positively influencing team affect. Four categories identified specific behaviors enacted during the process of contributing to group efforts (e.g. actively including team members, participating verbally, participating physically, and listening to others' ideas and perspectives). One category addressed the relative contributions of group members in relation to their abilities and included the concept of 'equitable' contributions in their definition. The final category specified that contributing to group efforts did not require speaking in the group. Even a greater number of definition categories (12) emerged from teachers' responses about what it means to 'do something nice for somebody.'

The responses that teachers provided to define each of these three behaviors are rooted in their training and professional development experiences, personal beliefs, and daily engagement

with students. For each behavior, teachers referenced specific training or curriculum. When defining what it means to ‘pay attention’, for example, teachers referenced SLANT, an acronym that summarizes specific tasks to support paying attention (e.g. Sit up straight, Lean forward, Ask questions, Nod head, Take notes). When defining what it means to ‘do something nice for somebody,’ teachers mentioned the book entitled, *Have You Filled a Bucket Today*, by Carol McCloud.

**Retrieval.** The second task in the cognitive response model involves retrieving information from long term memory that is relevant to the prompt (Tourangeau, 2003). Teachers were asked how they recall information about student behavior in the context of responding to rating scale prompts. For each of the behaviors queried, teachers were asked the following question: “In recalling how frequently a student pays attention (e.g. does something nice for somebody, contributes to group efforts), do you think specifically about recent experiences, do you average together experiences you’ve had over a number of weeks, do you think of a particular event, or do you think about it in some other way?”

Most responses reflected a tension between general perceptions of a student and a more methodical focus on specific events. Teacher responses that favored general perceptions described looking at the “big whole” (Interview #2) of a student, their “cumulative [experience] from the beginning of the year” (Interview #15), and/ or their ability to “think about the student in general” (Interview #5). One participant summarized this top-down approach,

I guess it's just an aggregate of everything. I just picture them in my head in the classroom and what are they most likely doing (Interview #8).

Another teacher described a similar method,

I guess I think about it more about an average...I might take one situation as an example of what I mostly see, versus picking one particular thing and saying that that is sort of exemplifies everything” (Interview #6).

Many teachers who characterized their process as reliant on general perceptions and opinions, reported considering the totality of student behaviors over time to form judgments. One participant noted, “It is really how I feel the student pays attention overall...It’s a very general teacher opinion” (Interview #10). In some instances, teachers described recall processes that did not include the active sampling or retrieval of specific experiences from long term memory; instead these informants reported calling to mind global impressions of a student.

The reported length of time teachers referenced when summarizing their perceptions varied widely. Among teachers using a top-down recall process, teachers reported sampling periods ranging from “recent weeks” (Interviews #4, #5, #9), to the entire year (Interview #11, Interview #13, Interview #15). One teacher noted that she compares her students to students she has had over the course of her career (Interview #14). Another participant reported using different time periods for different groups of students within the same classroom. This rater stated that he would sample a longer period, or ‘back story’, for kids who had received behavioral interventions than he would for students who had not received additional support in his classroom. He stated,

For...kids who are on tier 2 and tier 3 kinds of programs, I definitely think more back story for them, just because I've been working with them on it much longer. If you were to say, ‘Tell me about this kid's attention.’ I think I would most likely think about the last week or so (Interview #12).

Thus, teachers reported referencing different time intervals when retrieving information about their students. In some instances, the time interval considered when evaluating students ran counter to the directions provided on the DESSA-Mini, instructing informants to base ratings on observations from the past four weeks. One teacher challenged the appropriateness of relying on a prescribed time for evaluating behavior,

I would ignore that. I am sorry, but I would...I would totally ignore that past 4 weeks. Cause the last 4 weeks, they could have gone to Disney World and it doesn't matter to me. I am going to rate them based on what I really think about what that child needs and I would totally ignore that part of the question (Interview #6).

She reported that she makes her own determination about what period of time to consider when rating her students. In sum, teacher reported rating students on behaviors exhibited over different periods that were personally determined by individual teachers.

The top-down pattern described by some participants is consistent with assessment practices enacted by these informants as they rated students using a think-aloud process. Teachers' descriptions of their cognitive processes generally aligned with the rating practices observed when teachers rated their students aloud. All the teachers whose response patterns were reliant on impulse and general impressions, for example, reported utilizing top-down processes on which they based their ratings on general impressions of students that were sometimes substantiated, but not dependent on, recollections of specific events.

In contrast to the teachers reliant on impulse and general impressions, teachers reliant on reflective thought and specific examples (and to a lesser those reliant on reflective thought and summative experiences) were more likely to cite the importance of retrieving specific events and recent experiences from memory when recalling information about their students. These participants stressed the utility of calling to mind specific experiences when rating students. One teacher noted,

I guess also when explaining about it, it helps to think of a situation as well and I think it's like average with some of those little examples pieced in (Interview #4).

Another teacher shared,

I would first think about a particular event based off the questions you've already asked me; I'm thinking about some specific things that happened most recently. Then, that will lead to a more bigger picture (Interview #13).

A significant difference between teachers' cognitive processes based on general impressions and those reliant on specific events is the emphasis teachers placed on direct observations of student behavior. The raters who relied on specific events retrieved information from long term memory to bring actual observations of students to mind. Their ratings were dependent on recollections of actual events, rather than reliant on inferences regarding the frequency with which a student exhibited a behavior based on character, aptitude, or personality. One participant explained,

It would kind of need to be observable for me in some way... like the first kiddo I talked about, I see her really doing that [something nice for somebody] a lot, what I consider that to be. Doing those things leads to a number three and so that's why I rated her highly and the difference between her and the second kiddo I talked about is really more about observability. It is not that the second kid is not a super nice girl, right? Like anybody would be proud to have her as their daughter (Interview #4).

Teachers reported that the memories retrieved during the recall process were influenced by recency effects, by experiences that were out of the ordinary, or emotionally salient events. One teacher explained that more recent events are more easily recalled than more distant experiences. She explained,

I would say I kind of average... [student behavior] generally and I would probably say over recent weeks because there's that primacy or recency effect. It's hard to remember what they were like in November, you know (Interview #4)?

Another participant also recognized her tendency to consider events in the last few weeks,

I think about it probably the last few weeks, or specific things that we've been doing that are coming to mind (Interview #14).

It stands to reason that memories that are more accessible to the informant may be more consistently recalled than memories that take more effort to access.

Memories of events that are emotionally charged may be more salient and accessible than events with less emotional valence. This well documented finding in the academic literature, is described in teachers' own experiences when recalling memories of student behavior. Teachers

noted that “large events” (Interview #9) or a “significant event” (Interview #12) can sometimes affect teachers’ perceptions of a student, and emotionally charged events are more memorable and frequently come to mind when rating a student. One teacher explained,

Sometimes as a teacher, you become emotionally invested in an event. It frustrated you or it made you excited, and that sticks in your mind. Therefore, that then clouds over sometimes a rating of a child (Interview #10).

Another participant anxiously relayed his experiences with a significant event,

I was thinking about the field trip we were on where this kiddo was just kind of off his rocker and manic to a point that I've never seen before. He was making some unsafe choices by the stream and just it was like he couldn't pay attention to me or his chaperone or the other teachers. That was significant in my brain (Interview #12).

Each of these teachers explained that these memorable events quickly came to mind when rating the child involved. In sum, when reflecting on their own cognitive response to the rating scale measure, teachers reported differences in the types of information they recalled. Many respondents explained that recent, unexpected, and emotionally charged events can be especially salient during the recall process. The next section describes how teachers reconcile multiple memories to form judgments and select rating scale responses.

***Judgment and response selection.*** Teachers utilized different heuristics to form judgements about the frequency of three student behaviors: ‘paying attention,’ ‘doing something nice for somebody,’ and ‘contributing to group efforts.’ Teacher responses for how they rate the frequency of paying attention are presented to illustrate the wide variation in teachers’ teachers’ cognitive processes when formulating judgments and selecting responses to describe student behavior.

When explaining how they distinguish frequency ratings for paying attention, participants based their judgments on the following 6 approaches:

- Average time student attends to instruction (Interviews #7, 8, 15)
- Count the subjects in which the student attends to instruction (Interview #2, 10)
- Number of reminders to redirect the student's attention (Interview #1, 9, 13)
- Assessment of student's effort (Interview #14)
- Ability to independently complete work following instruction (Interview #6, 12)
- Impression of awareness or tendency to attend (Interviews #3, 5, 6, 10)

Among the quantifiable approaches, teachers reported personal cut points associated with response categories. One teacher explained she would rate a student 'frequently' for paying attention if the child attended to instruction 'about 75% of the time,'

I would say frequently is about 75% of the time to me. The occasionally means maybe twice a day need a reminder and frequently paying attention means it's almost not on my radar that they're going to need help. The very frequently, they're pretty much independent about paying attention and staying focused (Interview #7).

One of her colleagues, working in the same school, held a higher standard for receiving a 'frequently' rating. In her class, students must attend 80% of the time to receive a 'frequently rating'. She reported,

Frequently sounds more than 50 percent. It sounds like frequently is 80 percent of the time they're paying attention. There's like a big disparity between 40 percent and 80 percent. There's a giant range. I think I would tend towards frequently but then I have great kids (Interview #8).

In her explanation, this rater suggested that her tendency would be to rate her students using the 'frequently' rating because she has 'great kids'. Her statement reflected a tension that was evident in many teachers' explanations. While teachers often cited some way to quantify their recollected experiences, many relied on general impressions of students rather than quantifying observations. One teacher's response was particularly reflective of this tension. Initially, she explained that she determined her frequency ratings of paying attention by counting the numbers of times that a student was paying attention.

Yeah, I would say it would be the number of instances of paying attention. I do it by count. What percentage of the time are they you know (Interview #5)?

When asked to explain how she arrived at the percentage, she amended her response to explain that she also relied on her general knowledge of students.

Just like any human being...like if you are a mother, you know. I only have one natural daughter, but. You get to know them so well that you are with them every day. It's not hard. Even though there's 28 of them, you really do get to know them and you get to know their moods. It's easy because we spend so much time together so you know if they are grumpy, you know if they are tired. You just know (Interview #5).

Many teachers' responses reflected emerging ideas about the rating scale categories, and how these response choices describe student behavior. Teacher's responses were often meandering and lacked clarity that would be expected when a practice has become mastered. One teacher let out an exasperated sigh and shared, "These aren't easy questions...It's really getting me to think" (Interview #12). Another teacher offered feedback about the interview process:

Well basically, ...I am really glad that we did this, because I have not really thought about this. You know it's just one more thing that we've had to do (Interview #2).

Teachers' understanding of the measure and the consistency with which they assigned frequency ratings was still developing. There was wide variability across the 15 participants on how they judged the frequency of student behavior and selected responses to best describe those judgments.

Teachers' emerging ways of thinking about frequency ratings were also evident in the variable ways in which teachers judged the frequency of 'contributing to group efforts' and 'doing something nice for somebody.' Among the participating teachers, the frequency of contributing to group efforts was judged by considering the following: a student's 'eagerness' (Interview #4); her general 'leadership abilities' (Interview #7); and her 'body language' (Interview #14). Similarly, doing 'something nice for somebody' was evaluated in myriad of ways ranging from how often a student performed acts of kindness (Interview #1) to how frequently a student bickered or argued with other students (Interview #12). These differences in

the techniques that teachers used to judge the frequency of student behaviors revealed--that in the absence of clear guidance on how to think about the frequency—teachers are constructing their own methodologies when establishing personal assessment practices. Thus, interpreting teachers' ratings as measures that were performed using standardized procedures may be inappropriate and result in erroneous interpretations about students' abilities.

The final phase of the qualitative analysis shifted from an investigation of the cognitive tasks that teachers enact when completing rating scale measures, to explore the content that populates teachers' thoughts when rating students' social and emotional competencies. This wave of analysis focused on the child, classroom, and school variables included in the quantitative analysis to explain the how these factors influenced teacher ratings. Individual informant variables were not included in the quantitative analysis and thus are not explored as part of this study.

**Rater Effects. *Child factors-Age.*** Quantitative findings showed that teachers were likely to rate the social and emotional competencies of older students in their classrooms more favorably than those of younger classmates. For every additional month in age at the time of rating, relative to the average age of children within a classroom, students' social emotional scores increased by an average of 0.14 t-score points. While a student's advanced age relative to his classmates significantly predicted teacher ratings, the size of the increase was small. For a child 6 months older than the class average age, for example, we would expect a score increase of 0.84, less than a full t-score point.

During the think-aloud, none of the teachers referenced the relative age of a student in comparison to peers. Teachers made 10 references to age, maturity and/or development when justifying their ratings of students, and referenced development an additional 20 times in when

responding to open-ended questions about rating students' social and emotional skills.

References to maturity were used to describe students with strong social and emotional skills.

One teacher, portrayed a child who had transferred to her school during the current academic year,

This student is...when I read accept responsibility for what he/ she did. I immediately think. She is emotionally mature. I think about IQ and EQ and someone can have a high IQ and not a high EQ. This is someone who has both a high IQ and EQ. Umm so she entered [school name], very late, most recently in the year actually which is very difficult for any student to do and she made the transition seamlessly. She was confident from the very first day. She was able to get up in front of the class and introduce herself. Umm she's composed. She is clearly mature (Interview #5).

In describing the student, the teacher equated high emotional intelligence to maturity. In contrast, teachers referencing 'childish' or 'immature' (Interview #11) behaviors used these terms to describe undesirable social and emotional behaviors.

Several teachers attributed students' undesirable behaviors to their current stage of development. They used phrases like "[he is] only 8 years old" (Interview #13), "because he's a child" (Interview #11), or "for this age group" (Interview #1) when justifying why a child was not exhibiting pro-social behaviors. One teacher explained,

He's in that narcissistic stage, still, at this point, developmentally. I think he's developmentally a little behind social/emotionally. He wants to do nice things for people, he's just not that often aware of what needs to be done or what needs people have at that moment in time (Interview #14).

The teacher appeared to excuse the child's lack of regard for others and argued that the student's behavior reflected his lack of understanding, rather than a conscious choice or value to attend to his own needs.

Many of the participating teachers explicitly spoke with optimism about the malleability of students' social and emotional skills (Interview #7) and how these skills improved as students grew older. One teacher explained,

[Students] are growing and they are learning and they are changing constantly and they are maturing. I see such a difference between the beginning of the year and right now (Interview #1).

Another teacher described her experiences in teaching third grade students and contrasted her experience teaching younger students. She remarked on the difference in students' abilities to self-reflect on their behaviors. She shared,

I think this age in 3rd grade, that's the age where they can really start doing that forward, thinking that's self-reflecting. But I've taught 2nd and 1st before and I don't know if that's as easy for sixth and seven year olds to grasp. But I think in 3rd grade, it's been really interesting to see there's a big shift in self-awareness (Interview #13).

In contrast to the teachers who reasoned that students were not exhibiting skills because their developmental progression suggesting an understanding of typical developmental expectations, one third of the teachers (Interviews #2, 3, 4, 5, 6) expressly stated they lacked knowledge about what behaviors they should expect from their students. One teacher, for example, stated,

I don't know exactly what is 8, 9-year-old, what the typical 8, 9-year-old emotional... like how many times should they be able to manage this situation or how often should they be able to do this or that. I don't know the answer to this (Interview #6).

Teachers lacking an understanding of behaviors that are within normal limits for an age or developmental period, may find it difficult to evaluate those behaviors in her students.

One very experienced teacher expressed her grave concern that teachers lack knowledge about the skills that they are expected to evaluate.

We don't even know what it is we are rating, we don't know what an expectation is, these are developing human beings. And we don't know what would be developmentally appropriate at any stage, or if it's not an age appropriateness then what is appropriateness, are there stages of social emotional awareness that go beyond age? It's something that is not understood in terms of assessment and evaluation, in my opinion (Interview #3).

Teachers are thinking about the development of students' skills as they completed student ratings and shared their beliefs that these skills are malleable and change over time. Many also expressed serious concerns about their lack of knowledge of children's social and emotional

development. There is no evidence in the current study to suggest that teachers were thinking about a child's age relative to his classmates when rating a student. Perceptions of maturity and immaturity were associated with strong and/or weak social and emotional skills; however, teachers did not appear to think about the relative ages of children in their classroom when rating students. Thus, allowances are not being made for younger students, nor are there higher expectations for students who have had more time to develop skills.

*Gender.* In the current study, gender significantly predicted teacher ratings of students' social and emotional competency. On average, teachers rated females 5.46 points higher than males. The difference in scores between males and females were even greater at Title 1 schools, where teachers rated females on average 6.16 points higher than males. Gender differences were also observed in the ratio of female to male students that teachers selected to represent the strength (13:2), typical (8:7), and need ranges (6:9) of social and emotional competency. Despite the sizeable effect of being female, teachers made only 5 direct references to gender when rating students. Each of these references contained stereotypical representations of girls and boys.

Sometimes, I feel like some of my girls are a bit more demure and soft-spoken (Interview #14).

It means a lot to me...that a little boy is actually listening to me. The girls usually do, but the boys are usually off in their own world (Interview #15).

In 3 of the 5 instances, teachers recognized and checked their biases. One teacher, for example, questioned his selection of a boy as representative of a child within the need range of social and emotional development.

My first choice was going to be a male, but then I'm thinking males are probably, I don't know, we see them as being more behavior problems as a society. You know? I'm wondering if I want to bring up that kiddo or if I want to think about another female. I don't know. What do you think (Interview # 12)?

He acknowledged that he had planned on selecting a male student as a child in the need range, and questioned whether he should counter this bias by selecting a female instead. In another instance, a female teacher wrestled with her biases and beliefs about gender differences in certain cultures. She contrasted girls from Hispanic and East Indian racial and ethnic groups.

My Hispanic girls are usually a little bit less likely to take the lead, and take charge, and go forward in something, and speak their mind. I find ... these are massive generalizations, but it's just the generalizations based on the data I've collected in my head. I do find that my Indian families are much more driven. They're a little more serious, my Indian girls are, a little more serious I think. Whereas oftentimes the boys, I've found, from that culture are much sillier. They have a little bit more freedom to make mistakes. I don't feel like I see that all across the board (Interview #14)?

This educator acknowledged that her beliefs are based on ‘massive generalizations.’ She appeared to recognize the potential problem with such broad sweeping generalizations about students, but moved to justify that her beliefs were based on ‘data’ that she had collected. In sum, many teachers held stereotypical representations for male and female students. Some teachers explicitly recognized their biases and attempted to check how these biases may be impacting their ratings. In some instances, this recognition resulted in revised comments, and in other instances, the teacher justified the accuracy of her bias.

*Eligibility for support services.* The quantitative investigation tested whether a student’s eligibility for four support services teachers’ ratings of social and emotional competency. Eligibility status for special education services, disability accommodations, and English language learning services, and gifted education curriculum were entered as child level variables in the hierarchical model. The data showed that when controlling for classroom and school membership, eligibility for special education, disability accommodations, and/or English language learning services significantly predicted teacher ratings of students’ social emotional competency. When holding all other variables constant, students eligible for special education

services scored 8.91 t-score points below ineligible children. Students eligible for disability accommodations, on average scored 5.36 t-score points below ineligible peers, and students eligible for English language services scored 3.58 t-score points below ineligible students when all other predictors were held constant. In contrast, eligibility for gifted education had no significant effect on teacher ratings. There were no significant differences between students enrolled in gifted classrooms and those receiving general education curriculum.

The relationship between eligibility for support services and teacher ratings of social and emotional competency was also observed in teacher selections of students exemplifying the strength, typical, and need ranges of social and emotional competency. Among the 15 students selected to represent the need range, four students were eligible for English language services, one child was eligible for special education services, and one student was eligible for disability accommodations. These numbers are roughly commensurate with the proportion of students eligible for these services in participating classrooms. These proportions; however, were negatively skewed among children selected to represent the strength range. Among the students selected as having strong social and emotional skills, none were eligible for special education services, or disability accommodations. Only two of the students selected to represent the strength range were eligible for English language learning services.

*English language learning.* In total, teachers rated 9 children who receive ELL services. Each of these ratings was analyzed in its entirety to investigate the ways in which teachers think about language learning abilities when evaluating social and emotional competencies. Only two participants explicitly mentioned language as an influence on skill performance. One teacher described the strengths of a child who had recently immigrated to the United States.

If someone needs to understand something, and even with her limited English, her cognitive understanding is pretty high, she is able to kindly explain things, or make them more clear. And she listens to other students at times when they are not able to clearly ask things or are confused about something, and she can, she can articulate it for them (Interview #3).

The teacher used the phrase ‘even with her limited English’ to preface her description of the student’s strengths and to emphasize her accomplishment in helping others. Her favorable description of the student’s behavior recognized that learning a language, in addition to academic content, placed greater demands on the child than native speakers. This child’s status as an English language learner was acknowledged when the teacher evaluated her social and emotional skills.

In the second example, a Latina teacher—who had been a language learner herself—empathetically described the impact of being a language learner on her student’s confidence, behavior, and work avoidance. She described this child as carrying a ‘heavy backpack’ containing the burdens of developing English skills, poverty, and non-White ethnicity. She compassionately described his feelings of vulnerability and inadequacy and explained that these emotions contributed to frustration, acting out behaviors, and task avoidance. When rating this boy, she explained,

You know, he'd never felt successful prior to this year, never. You can see it. The start of the year, you know, bad language. It's that frustration. ‘I'm failing. I don't get this. They know that I don't get this.’ It's a lot of task avoidance, because he's vulnerable. ‘If I try, I'll fail. Then what?’...It's hard to invest when you feel so insecure about something (Interview #9).

This participant advocated for educators to push themselves to help English language learning students to speak more often in class so that they improve their listening and speaking skills. She implied that developing these skills will help to relieve their burden. She reported consulting with English language learning specialists to advocate on students’ behalf.

We've talked a lot with our ELL facilitators just that it's that speaking, the listening and speaking and just sharing in class. I think that that hasn't necessarily been targeted sometimes with kids, and I have seen it. With me I had to really push myself to say, "Okay, guys. We're going to these numbered sticks," because I've got two students who will just lead the way, which is great, but what about the rest of us (Interview #9)?

Each of these teachers' comments revealed their beliefs about the interconnectedness of language learning skills and the performance of social and emotional skills in the classroom. In the first example, the student was rated in the strength range for social and emotional competency. In the latter example, the child was rated in the need range. These examples suggested that for these teachers, the student's status as a learner of English helped to explain the behaviors exhibited.

In contrast, teachers' ratings of the other 7 English language learning students made no mention of the child's developing language skills. These students were described as "easily frustrated" (Interview #4), "behind" (Interview #11), "distractible" (Interview #2), and "inattentive" (Interview #14). Teachers characterized these students as unable to work independently (Interview #2, #4, #11) and dependent on teacher support. None of these teachers explicitly acknowledged the possibility that English language skills were related to students' abilities to perform certain skills. When rating a recent Chinese immigrant, for example, one teacher conveyed her concern that something may be physically wrong with the student. In running through her thoughts about what may have interfered with the student's ability to accept responsibility, she did not explicitly consider his language learning capabilities. She explained,

I at one point was worried about him if everything was clear in his mind. I mean, I was concerned do you think he had a tumor? Do you think he has a vision problem? Does he have a hearing issue that he's not hearing me? Does he just not get it? Because in the beginning of the year, he rarely accepted responsibility and just looked at me like, "What are you going to do now? What are you going to do" (Interview #11)?

As this participant weighed a range of possible explanations for the student's inability to accept responsibility, she did not note the possibility that he may not have understood what was being

asked of him. Instead, she suggested that the young boy's skill deficit could be explained by his physical condition (e.g. vision problems, hearing issue, etc.), rather than his temporary circumstance as a recent immigrant trying to adapt to a new language and culture.

Teachers' ratings of English language learners' social and emotional skills suggest that most teachers may not be accounting for the impact that language learning has on related skills. Many informants are appraising the observed inattentive, hesitant, dependent, and frustrated behaviors in these students as social and emotional skill deficits unrelated to language development. It is reasonable to expect that a teacher's appraisal regarding the reasons for a child's behavior informs the ways in which she responds to the behavior. A teacher who recognized that a child's reluctance to verbally participate in group discussions may be explained by the student's lack of confidence in his English-speaking skills, would likely respond differently than a teacher who interpreted the student's behavior as disinterested or unmotivated. In one case, the educator may support the student by scaffolding his verbal participation using GLAD strategies to frame verbal participation (Interview #9), in the latter case, the teacher may redirect the student by requesting his attention, or imploring him to increase his effort. These interdependent cycles of student behavior and teacher response--informed by teacher appraisal--may set the direction and pace of a students' future social and emotional development.

*Special education and disability accommodations.* In contrast to many teachers' lack of reference to language abilities when rating English language learning students' social and emotional skills, teachers' ratings of students with documented disabilities, typically referenced the disability when evaluating the student's social and emotional behaviors. Two teachers rated students eligible for special education, and one teacher rated a student receiving disability accommodations. Two of the teachers explained their students' behaviors in general and

attention difficulties, specifically as reflective of their Attention Deficit Hyperactivity Disorder (ADHD) diagnoses. One teacher described his student's poor attention as the manifestation of his ADHD.

...And pay attention, not all. He has an IEP, he has attention deficit but we know there's a lot other stuff going on there...He doesn't pay any attention. He's like on the floor rolling around, playing with things in his hands. He's probably about 10% on task in class. He's in the resource room for 2-1/2 hours a day, so mostly he's pretty intensely, but he needs more (Interview #6).

This teacher acknowledged how difficult paying attention is for his student and advocated for additional support services. She understood the student's inattention to be a manifestation of his disability. She described him as having difficulty accessing the curriculum because of his heightened awareness of non-instructional classroom activity.

Pay attention. I would say rarely. Part of that is both his ADHD...Part of it is, is he not able to access the curriculum? Is he not understanding things? Whole group is very difficult for him, because he's very aware are what other people are doing. Paying attention, I would definitely say rarely. Small group, it's a little different, but it still requires, even with a group of 4 or 5 kids, a couple of callbacks (Interview #14).

Each of these teachers recognized the students' abilities to pay attention were attributable to diagnosed disabilities rather than a lack of effort or interest.

*Race and ethnic heritage.* In the current investigation, a child's race and/or ethnic heritage significantly predicted teachers' ratings of social and emotional competency. When holding constant other variables, and controlling for classroom and school membership, Asian students were rated on average 0.70 t-score points higher than their White classmates. African-American students and Latino students were rated 4.46 t-score points and 1.19 t-score points lower than White students, respectively. As previously reported, Latino and African-American students were underrepresented among students that teachers selected to represent the strength range of social and emotional competency. A higher than expected proportion of Asian students

(based on quantitative findings) was selected to represent the need range. These proportions were likely influenced by the racial and ethnic make-up of students within participating classrooms. One third of participating classrooms, for example, had no African-American students and in 4 classrooms, Asian students comprised more than half the class.

Four significant findings regarding a child's race and ethnic heritage emerged from teacher's ratings and their responses to questions inquiring how they think about race/ethnic heritage when evaluating students' social and emotional competencies.

First, teachers working across school contexts hold stereotypical beliefs about students based on their race or ethnic heritage. Participants working in both Title 1 and Non-Title 1 schools referenced students' race/ethnic heritage when rating student behaviors. Frequently, informants referenced a student's race or ethnicity when explaining behaviors that interfered with the child's performance in the classroom. When sharing information about a 'stubborn' child resistant to adult support, for example, the teacher explained that the young girl behaves that way because she is Puerto Rican. She offered, "Puerto Rican women are very strong willed and very proud women." The educator elaborated, "I understand [her behavior], because that's how her mom is. That's how her culture is" (Interview #1).

In another instance, a teacher attributed 'aggressive' and domineering behavior to a child's life experience as a Muslim male. In describing his 'maladaptive' tendencies to dominate group work and his 'disrespect' for women, the teacher offered,

[They're]... a very traditional Muslim family and his mom shared...that in their home, his dad is very much the dominant person so it's taught to him that you are that way (Interview #4).

This teacher wrestled with how to address the conflicting expectations placed on the child, and acknowledged that her behavioral values acquired through her own upbringing and personal experiences were at odds with those of her student's family.

What I observe him doing in groups, for example, I wouldn't consider productive, but if his dad were to come and see what he's doing in group, he would give him 5 stars (Interview #4).

She lamented her lack of knowledge in how best to guide the young boy.

I need to think about that. So, these things that I would consider maybe maladaptive for him, this kiddo is just trying to figure it out. He's 8 (Interview #4).

Many teachers, like the one here, attributed children's behaviors to their experiences as members of specific religious or racial and ethnic groups, in which behavioral norms differed from those of the teacher. The following comments were made when evaluating Latino, Asian, Chinese, and Indian students:

I am trying to think of his culture... but it's more of a Latino culture. He's a very nice boy, but he has like a testosterone problem and when he gets out to recess, he can be aggressive (Interview #5).

Asian students are far less likely than other ethnicities to share necessarily their thinking when it comes to this kind of issue. They are just naturally quieter I would say (Interview #5).

This year, I have five students who are new from China. Yeah, and so I know there's a different way that they look at adults, teachers especially (Interview #11).

I do find that my Indian families are much more driven (Interview #14).

Teachers' comments suggested that they are working hard to make sense of their students' experiences and to understand factors that may be influencing behaviors that differ from what is expected in the classroom setting.

Second, most teachers recognized their biases. Many teachers expressed grave concern that they were biased in their perspectives about student behavior, and seriously questioned the fairness in evaluating students from other cultures based on their own cultural perspective.

I know that I have my own biases that I ... I would tell you I don't take any of that into ... I think that's a negative as well that I need to be more aware of how different cultures and different ethnicities how they do regulate their emotions and ... That's the different than the way I regulate my emotions. I'm not aware of those things. I just assume that everyone does it the same way that [I do] (Interview #6).

I am a White person teaching black kids... I'm a person whose had... what is the word that I am looking for... privilege my whole life, so I guess I judge these based on a White privileged perspective even though I could say I don't pay attention to any of those factors when I am evaluating a student, but I know that that stuff is there (Interview #2).

We talk a lot about how our perspectives and our view of what American culture is acceptable and encouraged and how they may not be fair to judge students whose cultures are very different from that (Interview #7).

I am dominant White culture, right? Like I was raised in this culture. I am constantly feeling like I need to check myself. Am I expecting a kiddo to be a certain way because that's what dominant White culture expects? Right? (Interview #4)

In sharing these concerns, teachers squirmed in their seats, sighed, and expressed distress in not knowing how to reconcile their biases with their evaluation of student behavior. They expressed the desire to do what is right by their students, and yet shared their uncertainty about how to ensure fairness in their assessment practices. One teacher summarized, "I don't want to judge them based on a White idealized world" (Interview #1).

Third, teachers hold mixed views on whether ratings should be adjusted based on consideration for a child's race/ethnic heritage. While most teachers recognized their personal biases and the influence these biases placed on expectations of student behavior, more than half of these teachers (8) reported that they did not consider a child's race or ethnic heritage when

completing student ratings. In response to a question on how race or cultural heritage informs, or does not inform a teacher's thinking, one participant responded,

Well I try not to make it a factor in these questions because, but I mean...but those factors are there. They are present...It's like systematic racism like we... (pause). I feel like. I feel like I just can't answer this question (Interview #2).

Her struggle to engage in fair assessment practices was evident as she acknowledged her biases, but lacked direction in how to address these biases when rating students. Others emphatically claimed, "I try not to let it influence me" (Interview #11) and "It doesn't inform my thinking" (Interview #15). In explaining, why she does not factor race into her teacher ratings, one educator explicitly cited her lack of preparation in knowing how to deal with issues of diversity. She conceded,

I don't have the background. I am a little bit aware. I know that some of my kids are coming from backgrounds where they completely defer to adults and they need to ask about everything. That could be seen as not taking responsibility or not being problem solvers. I don't know how much we need to be talking about that... I don't have ethnic training or cultural training (Interview #8).

This teacher acknowledged her lack of training and communicated that she felt unprepared to account for race, or ethnic heritage, when evaluating students' social and emotional skills. In a separate interview, a colleague expressed a similar lack of confidence in her ability to fairly address the effects of race in her evaluations. She reported that she has a difficult time determining the racial and ethnic backgrounds of her students. She said that she thinks of students by their names and not by their appearance. She offered, "I don't seem to notice race well enough" (Interview #10).

In contrast to the teachers who reported not accounting for race, a smaller number of teachers (4) claimed they intentionally factor a child's race/ethnic heritage into their ratings. One teacher explained that she adjusted her ratings for non-White students. She shared,

I don't want to judge them based on a White idealized world. Yea, I do think it plays a very important...I factor in it (Interview #1).

Another teacher argued that regardless of teachers' reports, her ratings, and all other teachers' ratings are impacted by a child's race:

That child's interests, the child's personality, the experiences you have with the child, all the observations you make throughout the day, so to say that ethnicity does not have an influence would not be true, because everything about the child has an influence (Interview #3).

The remaining teachers (3) reported that they either take race and ethnic heritage into account 'somewhat' or they are unsure of how it influences their ratings.

Fourth, none of the teachers offered an approach for how to address racial and ethnic differences in teacher ratings of students' social and emotional competency. Among teachers who reported adjusting ratings to account for race it is unclear how they would adjust and under what circumstances. Would teachers reference children of the same race and gender when completing ratings? What information will they use to make determinations about the norms, expectations, and values of specific races/ ethnic groups? How will these ratings be modified or adjusted for children who live in multi-cultural homes? Teachers asked for training to help them know how to evaluate their students fairly and reliably. One teachers, implored, "The teachers definitely need training, because it's a new area for us" (Interview #5).

*Contextual factors.* In the present study, the classroom and school variables evaluated did not significantly predict teacher ratings of students' social and emotional learning. When controlling for classroom and school membership, class size and the percentage of students from historically disadvantaged racial and ethnic groups did not significantly predict outcome measure scores. The number of students enrolled in a school, the percentage of students eligible for meal subsidies, and the average number of years of teaching experience also failed to account for

significant differences in teacher ratings of students' social and emotional competency. In the teacher interviews, only three teachers referenced the classroom environment when rating students' social and emotional skills. While teachers frequently described the influences of peers on individual students (33 coded references), very few respondents spoke about classroom climate as an influence on the performance of student skills, and none of the teachers referenced the school climate when rating students.

In each of the three instances in which raters referenced classroom climate, the teacher described the classroom impact on student's skills. In one instance, the teacher noted classroom conflict and the additional demands the context placed on a student.

There's been a lot of conflict in my room, which we only have 7 girls... There's been some pretty intense peer conflict that's occurring even now. She mostly stays out of it, but I don't think she takes the advice enough to apply to others and to really remove herself from that. I would say she can work on that (Interview #7).

These demands allowed the teacher to observe the student under atypical classroom conditions. It is unclear to what extent the teacher altered her expectations for student behavior. In recognition of the additional challenges placed on students, do the expectations for behavior remain constant or are they adjusted to account for the increased demands? In another example, the teacher praises a child for her ability to remain positive in a trying context. In describing a student in the strength range of social and emotional competency, she reported,

She's very patient, because I've got a lot of emotional needs in the class this year. She's very patient with, as best as she can, speak about positive things (Interview #14).

Contextual factors do not appear to play a significant role in teachers' thoughts as they evaluate students social and emotional competency. Even among the few teachers who explicitly acknowledged unusual classroom conditions, such as a significant gender imbalance, extreme

conflict, or great emotional needs, teachers were not intentionally considering students' performance of social and emotional skills within the broader classroom context.

## Chapter 6: Discussion

Social and emotional skills play a pivotal role in educational, career, and life outcomes. For the past 30 years, the research literature has consistently documented the impact of social and emotional skills on academic performance, interpersonal relationships, employability, and personal well-being. The Collaborative for Academic Social and Emotional Learning (CASEL), established in the 1990s, has taken the lead to coordinate the efforts of researchers, policy makers, and educators, to establish a research agenda to identify best practices to guide the implementation of social and emotional learning (SEL) programs in American schools. CASEL introduced a social and emotional learning framework consisting of 5 core competencies that has been used to set local, state, and federal educational policies governing the promotion of SEL programming in schools. The CASEL framework has also established learning targets for most universal SEL curricula implemented in schools today.

The SEL movement has made considerable gains in promoting the inclusion of SEL skill instruction in schools. Establishing complementary efficient, reliable, and unbiased assessment practices to evaluate the impact of instruction on students' skills has been more challenging. Although there is burgeoning research documenting the effectiveness of SEL programs in schools, little is known about the ways in which teachers evaluate students' social and emotional skills as part of their teaching practice. Interested stakeholders continue to debate what skills to measure, how to measure them, and if objective measurement of such complex skills is plausible, particularly among diverse groups of students. The current project addressed this gap in the SEL research literature.

Teacher rating scales are amongst widely used modalities for evaluating social and emotional skills in children (Denham, 2015; Denham et al., 2010; Merrell, 2001). They are

prized for their efficiency, low-cost, and ability evaluate many skills and competencies using a single measure. Teachers are well positioned to observe students' behaviors under a variety of setting demands across significant time spans. Yet, substantive questions remain surrounding low inter-rater reliability (Achenbach et al., 1987), teachers' thought processes when completing standardized rating scales (Duckworth & Yeager, 2015), and the appropriateness of using Likert-type scales to evaluate diverse groups of students (Heine et al., 2002; Peng, Nisbett, & Wong, 1997). To date, there have been no studies investigating the ways in which teachers interpret and respond to rating scale measures.

To address this gap in the research literature, the current sequential explanatory study investigated the assessment practices of 173 elementary classroom teachers in a minority-majority school district in the Pacific Northwest. Teachers rated 4005 grade 3-5 students (N = 4005) using a standardized rating scale (e.g. DESSA-Mini), per contractually mandated district assessment requirements. Hierarchical linear analysis techniques were employed to evaluate the relationships among child, classroom, and school variables and teachers' ratings. Subsequently, 15 teachers were purposively selected from the full sample to participate in semi-structured interviews designed to explore the ways in which educators' thoughts, experiences, and beliefs contributed to their ratings.

Quantitative results found that six child level variables significantly predicted teacher ratings. Classroom (e.g. class size, and % of historically disadvantage students) and school level covariates (e.g. average years of teaching experience, % free or reduced lunch, and school enrollment) did not significantly predict teacher ratings of students' social and emotional competencies. Consistent with the research literature, a child's gender (Winsler, Kim, & Richard, 2014) significantly predicted teacher ratings. Rating scale respondents reported that

girls displayed behaviors associated with positive social and emotional competency (e.g. paying attention, doing something nice for somebody, speaking about positive things) more frequently than their male classmates. This finding conforms to the research literature documenting developmentally earlier rates of social and emotional maturity in females (Naglieri et al., 2009), and the relative proclivity in males to exhibit externalizing behaviors more frequently than female age-mates (Epstein et al., 1998; S. Hinshaw, 1992; Powell, & Kamphaus, 1994; Wolcott & Williford, 2015). Externalizing behaviors are often visible, disruptive, and elicit emotional reactions within the classroom environment. These memorable experiences are likely more readily retrieved and salient when teachers recall prior observations of student behavior. The quantitative findings from the current study are also consonant with gender stereotyped thinking found in some teacher interviews. Interview participants described boys as active, busy, non-listeners (Interview #15), and girls as quiet, and soft-spoken (Interview #14).

A child's chronological age relative to peers also significantly predicted teachers' ratings. Concordant with previous literature, teachers rated older students within a classroom more favorably than younger classmates (Mashburn et al., 2006). This result may be explained by the expected improvements in students' ability to regulate their emotional states and related behaviors because of more life experience. In relation to younger classmates, older children were viewed as more socially and emotionally skilled than their younger classmates.

In the current study, a child's race or ethnic heritage also predicted teachers' ratings. In the overall sample, significant effects were observed for Asian, Black/African-American, and Latino/Hispanic students in comparison to White students (see Table 5.19). Additionally, racial/ethnic group effects were dependent on the school context in which the child was rated. Separate sub-sample analyses were conducted to evaluate the race/ethnic heritage effects in Title

1 schools and Non-Title 1 as Non-Title 1 schools. In Title 1 schools, identification as Asian or Black/African-American significantly predicted teachers' ratings, yet identification as a Latino student did predict teacher ratings (see Table 5.27). At Non-Title 1 schools, in contrast, being Black/African-American or Latino/Hispanic significantly predicted differences in teacher ratings in comparison to White students (see Table 5.26).

The moderating effect of the school context (e.g. Title 1 vs. Non-Title 1) suggests that the reference group of students that a child was compared against may have influenced teachers' judgments about the child. In the current investigation, Latino/Hispanic students comprised 37.13% of students in Title 1 schools. In contrast, they made up only 19.24% of the student population in Non-Title 1 schools. Latino students evaluated in school contexts in which a larger proportion of students identified as Latino, may have exhibited social and emotional skills that were judged to be 'typical' when compared to students sharing similar social and emotional norms from similar racial/ethnic backgrounds. These same students, however, may have been judged by other respondents as 'atypical' or in the 'need' range when compared to a majority of students who exhibit customs and norms that differ from the minority group.

The effects of reference group bias are well documented in the academic literature (Duckworth & Yeager, 2015; Heine et al., 2002; Kautz, Heckman, Diris, Weel, & Borghans, 2014; Peng et al., 1997; Tesser & Schwarz, 2008), and may be more pronounced in cross-cultural comparisons amongst groups prone to stereotype bias (Duckworth & Yeager, 2015). The use of Likert-type rating scales, like the DESSA-Mini, maybe especially problematic when the frames of reference against which individual students are judged, vary significantly and systematically across respondents using the same measure (Duckworth & Yeager, 2015).

Reference group bias effects may also occur when a student's cultural norms differ from those of the rater. In both Title 1 and Non-Title 1 school contexts, the effects of a child's race were consistent with stereotypical representations of the groups' relative achievement. Asian students were rated more positively than their White classmates, and Latino/Hispanic and Black/African-American students were rated less favorably. The prevalence of stereotype bias was also evident in teachers' think-aloud ratings of their students. In semi-structured interviews, teachers described Asian students as the model minority (see Petersen, 1966). They were characterized as intelligent (Interview #7), hard-working (Interview #1), driven (Interview #14), and quiet (Interview #5). In contrast, children from non-White non-Asian cultures—historically marginalized in schools-- were described as athletic (Interview #15), aggressive (Interview #4), stubborn (Interview #1), and passive (Interview #9).

Most teachers recognized their biases and explicitly expressed concern that they may be unfairly evaluating their students based on behavioral expectations acquired through life experiences that have been transacted in cultural contexts that differ from those of their students. Implicit in many teachers' comments is the recognition that each student's social and emotional development is shaped by the unique combinations of attitudes, behaviors, and values they observe in their daily lives. The social learning (Bandura, 1977b) and personal experiences of that have transpired over time in different home contexts (Bronfenbrenner & Morris, 1998), contribute to differences in the social norms that govern the ways in which students and teachers act and appraise behavior. Some teachers lamented their limited understanding of cultural differences and relayed their lack of training in knowing how to guide students to navigate the 'code switching' (Interview #4) required to uphold the behavioral norms at both home and school.

To date, the research on the racial/ethnic match between evaluator and student remains inconclusive. Some studies find no racial effects on teacher evaluations for non-White children (Conners, 1969; Gresham & Elliot, 1990; Mashburn et al., 2006). Other investigations find rater effects associated with the evaluator's racial/ethnic heritage. Pas and Bradshaw (2014), for example, found White teachers rated students of all racial/ethnic backgrounds more favorably than their Black/African-American colleagues. In contrast, some investigations have found that teachers rate students sharing their racial/ethnic heritage more favorably than teachers from racial/ethnic groups that differ from those of the student (Downey & Pribesh, 2004; Ehrenberg et al., 1995). The racial/ethnic identities of teachers in the current study were unavailable which precluded an analysis of the effect of racial congruence on teacher ratings.

A student's eligibility for support services significantly predicted teacher ratings of students' social and emotional competencies. Students eligible for special education, disability accommodations, and English language learning services, on average, received lower social and emotional competency ratings than their peers in both Title 1 and Non-Title 1 schools. Lower ratings were expected among students found eligible for services because of social and emotional disabilities and among students with conditions that compromise their ability to attend to instruction. In the semi-structured teacher interviews, participants rating students with attention difficulties, or emotional and behavioral disabilities, clearly attributed deficiencies in students' social and emotional behaviors to their disabilities. Teachers spoke compassionately about these students and the difficulties they experienced accessing curriculum, and forging positive relationships. Additional research is needed to disaggregate factors associated with special education services and disability accommodations to learn more about how specific learning disabilities (e.g. dyspraxia, dyslexia, or dysgraphia) impact social and emotional development.

Lower ratings were also observed among students eligible for English language learning services. These lower scores may reflect the challenges that teachers, peers and English language learners have in understanding one another during classroom interactions. Emerging language development may make it difficult for language learners to fully participate in the classroom environment and form relationships with other members of the classroom community. Getting to know these students and understanding their behaviors may require more effort from teachers and peers to understand and bridge the language and cultural differences. Language differences that compromise a child's ability to communicate with his schoolmates and teacher may also contribute to sustained stress places the student at greater risk for heightened emotionality or withdrawal behaviors that may interfere with school functioning.

There is a dearth of researcher investigating the relationship between second language learners and the development of social and emotional skills. Existing literature documents the negative impact of poor social and emotional skills on language acquisition (MacIntyre & Gardner, 1991; Winsler et al., 2014); however, additional research is needed to investigate how status as a language learner impacts the acquisition and performance of social and emotional skills at different points in development. In contrast to teachers' recognition that a child's capacity to focus impacted social and emotional development, teachers typically did not account for students' language differences when rating social and emotional skills. Moreover, teachers generally failed to address a child's language learning skills when evaluating the ability to attend to instruction, do something nice for somebody, or contribute to group efforts. Instead, many teachers attributed students' difficulties to factors unrelated to language development.

Teachers' explanations for student behaviors warrant our attention because their appraisals inform the ways in which they respond to students and intervene to support

development. A teacher attributing a student's lack of participation in group discussions to his challenges keeping pace with interpreting and responding to an unfamiliar language, will predictably respond to the child in a different manner than a teacher who attributes the student's reticence to disinterest, boredom, or a lack of motivation. Additional teacher education is required to ensure that those working with English language learning students understand factors that interfere with social and emotional skill performance so that they are equipped to accurately appraise student development in real-time and intervene to address the true barriers to skill performance.

In addition to evaluating the impact of child, classroom, and school predictors on teacher ratings, statistical analyses were used to partition the variance across the three-level model to determine the proportion of variability in student scores located at each level. In the full sample, 81.79% of the total variability in teacher ratings was accounted for by within-classroom differences. A smaller, but sizable proportion of variability was explained by between-classroom (15.60 %) and between-school (2.61%) differences (see Table 5.14). The proportion of total variability located at the between-classroom level was higher than expected. Assuming the students populating classrooms within the same school were reasonably similar in their social and emotional skill performance, we would have anticipated a higher proportion of the total variance in students' scores to be attributable to within-classroom differences, reflecting the range in students' social and emotional competencies within a classroom. While most variability in student scores was, in fact located at the within-classroom level, the proportion of variance at the between-classroom level exceeded anticipated results.

Separate analyses were also conducted segmenting the full sample by Title 1 school status. Interestingly, the proportion of variability explained by each level of the model differed

by school type. The within classroom variance among Title 1 schools comprised a greater portion of the variability than the full sample. In comparison to the full sample, an additional 4.79% of the total variability was explained by within-classroom factors. The between-classroom differences dropped to 13.31% and between-school differences were reduced to 0.25% (see Table 5.25). In contrast, a higher proportion of the total variability in Non-Title 1 ratings (16.60%) was located at the between classroom, or between teacher level with an even lower percentage attributable to between-school differences (0.11%) (See Table 5.16).

The larger than expected differences between classrooms suggest that teacher factors may be contributing to rater effects that warrant consideration. Differences in results for Title 1 and Non-Title 1 schools may indicate that the nature and magnitude of rater effects are moderated by the school type and the assessment practices of teachers working in those contexts. Teacher ratings in Title 1 schools contributed to greater variability across students within a classroom and smaller between-classroom differences. Accordingly, two grade 3 teachers working in the same building have classroom score profiles that are more similar than different. In comparison, between-classroom differences were in Non-Title 1 school contexts in which a higher proportion of the variance was located at the between-classroom level.

To explain these quantitative findings, semi-structured interviews were conducted to learn more about teachers' thought processes when completing student ratings. While the present study was based on teacher ratings using the DESSA-Mini, it would be reasonable to expect similar response patterns in teacher ratings using similarly structured measures. The CASM research literature, suggests that respondents engage in the same set of cognitive tasks (e.g. comprehension, retrieval, judgment, response selection) in a similar manner when

completing standardized survey measures, or rating scales. Yet, little is known about how teachers perform these tasks under ecologically valid conditions.

In the current study, teacher *think-aloud* ratings of their students made their thinking processes visible and explicit. Teachers verbalized thoughts revealed significant and unexpected differences in the ways in which teachers rated students' social and emotional competencies. Three response style prototypes emerged from respondents' ratings. These include: 1) responses reliant on impulse and general impressions, 2) responses reliant on reflective thought and summative experience, and 3) responses reliant on reflective thought and specific experiences.

Each of the three response styles was associated with substantive differences in 1) the amount of content teachers retrieved from memory to complete ratings, 2) the contrast in response selections across behaviors within individual assessments, 3) informants' confidence in the accuracy of their own ratings, d) and the congruence between teachers' comprehension of a behavior and the consistent application of her understanding to assessment practice (see Figure 5.1). Teachers reliant on an impulse and impressions response style took a top-down approach when evaluating students and typically referenced a child's personality traits, or character, when explaining ratings. Evaluations of individual students often reflected little or no variability in response selections across the 8 behaviors measured. "A great kid" for a teacher reliant on general impressions and impulse was consistently great, regardless of the behavior targeted by a prompt. He was great at paying attention and great at contributing to group events. In contrast, teachers reliant on reflective thought and specific experiences were more apt to relay specific experiences of individual students exhibiting the targeted behavior. Teachers reliant on reflective thought and specific experiences were more confident in the accuracy of their ratings and their evaluations were characterized by greater within-student variability.

Interestingly, teachers' response styles were related to the school contexts in which they worked. Eighty percent of the teachers reliant on reflective thought and specific experience worked in Title 1 schools. These teachers' quantitative ratings contained greater within-student variance across the 8 behaviors assessed. Accordingly, the predominant response style exhibited by Title 1 teachers may be more effective in distinguishing students' abilities across many behaviors in comparison to teachers reliant on general impulse and general impressions. This finding is consistent with the differences observed in the proportion of variance attributable to within and between-classroom differences observed at Title 1 and Non-Title 1 schools and the proportions of student ratings in each of these settings with no within-student variability in scores. All, but one of the teachers reliant on impulse and impressions and reflective thought and summative experiences worked in Non-Title 1 schools.

The cognitive response styles identified to describe teachers' response patterns were also associated with significant differences in how students were rated. This finding holds significance when interpreting teachers' ratings of students' social and emotional competencies. Typically, when interpreting individual student scores and classroom mean scores, stakeholders assume that students are rated using a 'standardized' assessment practice that is consistent across raters. The current investigation provided new evidence that teachers are not completing rating scale measures in a standard way and differences in teachers' response styles impacted the ways in which they assigned ratings to individual students, and the scores derived from these ratings.

Teachers' thought processes were also probed during the teacher interviews to explore teachers' comprehension of rating scale items, and the ways in which they retrieve information about students, form judgments, and select responses from available choices. Teachers' responses revealed substantive differences in teachers' conceptions of behavior, the ways in

which they thought about the frequency of student behaviors, how they formed judgments, and their selection of rating scale response options.

In many instances, one teacher's comprehension of a behavior and the ways in which she evaluated the frequency of that behavior were in direct conflict with the definitions and approaches described by other respondents. Teachers working in similar contexts tended to hold similar standards for student behavior. Participants working in Title 1 schools, for example, expressed more tolerance in behaviors associated with paying attention. These educators were less likely to require students to sit upright, nod, and have eyes on the speaker than teachers working in Non-Title 1 schools. One plausible explanation for these differences in teachers' standards for behavior is that educators working in Title 1 schools encountered classroom experiences that challenged their previously held standards and beliefs about student behavior. These experiences within the school contexts in which they practice contributed to revised expectations, standards, and beliefs that differ from teachers working in other contexts.

As previously discussed, teachers also retrieved different information when rating students. Some respondents relied on general impressions of students, and others emphasized the importance of considering 'observable' behaviors in their ratings. Teachers' response selections also differed across respondents. In some instance, respondents utilized a truncated range of response options, eliminating selections at the upper and/or lower end of the response range. In other instances, respondents assigned the same rating across all behaviors for every student in the classroom. Consequently, students enrolled in classrooms in which the teacher utilized a restricted range of responses did not have the same range of possible scores as a student enrolled in classroom in which the teacher utilized the full range of response options. In

sum, teachers used markedly different standards to evaluate student behaviors, and employed different frequency measures when applying these standards.

## **Chapter 7: Conclusion, Limitations, Implications and Future Research**

So, what's in a rating? This project leveraged the respective strengths of quantitative and qualitative methodologies to explore how teachers are engaging in the processes of evaluating students' social and emotional skills under real-world conditions. This study is the first to pair a quantitative analysis of teacher ratings--completed under ecologically-valid conditions--with a qualitative investigation of teachers' thought processes enacted when completing student ratings. Prior social and emotional measurement research has largely focused on the associations between standardized measures and various predictors. This project builds upon this body of research by introducing teacher voice to the academic literature to explain how practitioners interpret and respond to social and emotional rating scales when evaluating students' social and emotional competencies. The results from this mixed methods study revealed that many factors other than student behavior influence teacher ratings of students' social and emotional competencies.

First, teachers held different behavioral standards for students that are shaped by personal life experiences, training, and daily interactions with students. Second, teachers varied in the ways in which they were applying these standards when rating students. Some relied on methodical application of their standard to specific experiences with individual students. Others, took top-down approaches and applied ratings based on overall impressions of students, referencing factors that may be unrelated to social and emotional competency. Third, teachers relied on different interpretations of frequency responses and used different ranges in their responses when evaluating students. These personal choices to utilize a limited range of responses restricted the ratings obtained by students within their classrooms. Fourth, and possibly most importantly, teachers recognized that students' social and emotional skills are

shaped by the behaviors, values, attitudes, and skills of the adults, peers, and community members they observe in the contexts in which they live. They also acknowledged differences in expectations at home and at school that some children experience. Yet, educators have lack training in how to account for these differences in their ratings, if at all. In the absence of guidance and direction for how to address these differences, teachers are doing their best to make professional and fair decisions. For some teachers, that means applying different standards to children from different racial and ethnic groups, and for others it means not taking race, ethnicity, religion, or familial influences into account.

These findings cannot be ignored when interpreting students' scores on measures of social and emotional competency. What can be said about a student scoring in the strength, typical, or need range? Did the teacher base the child's score on a top-down assessment of the child because she is compliant? Does the score reflect the teachers' tendency to rate all students favorably? Does the score reliably say anything at all about the social and emotional behaviors that a child has exhibited in the classroom? The current project suggests that we cannot be sure. Currently, there is too much noise in teachers' ratings of social and emotional competency to trust that a score accurately describes a student's social and emotional competencies.

### **Limitations**

The present study is limited by its reliance on a single district for participants. It is unclear if the associations between teacher ratings and the predictors tested are representative of relationships that exist in broader population of teachers. Because of contractually negotiated privacy agreements for teachers in the participating district, there was limited access to individual teacher information (e.g. years of teaching, race, gender, performance review scores) that could have helped to describe the relationships between informants and their ratings.

Reliance on a single teacher measure of social and emotional competency also compromised the ability to look at patterns in teacher assessment practices across time or resulting from the use of different measurement instruments or techniques. Independent observations of students within teachers' classrooms would have also increased the rigor of the investigation, by allowing for triangulation of student scores across measurement modalities.

The qualitative investigation is limited by the number of participants and possible selection bias. There were only 15 teachers participating in the interviews and they all work in the same district. Each of the participants volunteered to participate in the study. While the sampling was purposeful to include teachers working in Title 1 and Non-Title school contexts with a range of experiences for maximum variation in the sample, there was likely some selection bias in favor of teachers who support the inclusion of social and emotional learning programs in schools.

### **Implications**

Notwithstanding the described limitations, the project suggests important practice implications.

**The use of teacher rating scales as sole measures of students' social and emotional competencies should be discontinued.** Despite the best practice recommendations that consistently discourage the use of a single measure or single informant approach to describe students' social and emotional skills, practice constraints frequently compromise best practice recommendations when completing universal assessments. Results from current project highlight the significant rater effects that threaten measurement validity and reliability. If teacher ratings are to be retained as a reasonable measure of students' skills, we must increase the consistency in teachers' comprehension and response processes when rating students.

**Teachers require additional education on social and emotional development.** Formal social and emotional skill instruction constitutes a new content area that teachers are increasingly expected to teach and to assess. Quality instruction and assessment require a theoretical foundation for the content being taught and the skills assessed. Teachers would benefit from professional development that increases their understanding of how social and emotional skills develop and the mechanisms that promote and interfere with growth. In contrast to the child development training that pre-service teachers currently receive--often divorced from application to classroom practice--future efforts must offer a bridge between theory and practice. It is incumbent upon teacher educators to ensure that teachers understand how a child's academic capabilities, cultural background, poverty, and status as an English language learner influence the social and emotional behaviors he exhibits in the classroom so that teacher appraisals and responses promote prosocial student development. Furthermore, educators must develop a working knowledge of how typical development unfolds and the expectancies for skill performance at different ages and under different conditions. Finally, teachers require diversity training to enhance their knowledge of how racial and ethnic differences influence student behaviors. Teachers are requesting direction for how check their biases and consider race and ethnic heritage to ensure fair and unbiased ratings of all students.

**Improved measures of social and emotional competency need to be developed.**

Teacher rating scales are amongst the most widely used modalities for assessing students' social and emotional skills. As previously noted, they are prized for their efficiency and ability to evaluate many behaviors in a single measure. Additional work must be done to develop measures that retain these benefits while addressing some of the pitfalls associated with them. The current use of frequency ratings in which a single response selection is consistently

favorable or unfavorable, fails to counteract response patterns among informants inclined to rate a student quickly with little regard to variation in behaviors. Such top-down response patterns could be minimized by intermittently reversing the direction of items. Instruments could also have built in response pattern checks that flag responses following a predictable pattern. The additional of this check would identify teachers, like those in this study, who rated all behaviors for every child using the same rating. The ratings for children within these classrooms are completely useless. Flagging these respondents provides an opportunity for discussion about their processes in rating students and further training.

Another approach for addressing response patterns is to alter the format of rating scales. Teachers could instead be provided with short vignettes of common classroom experiences (e.g. small reading group, whole class discussion, disagreements in class) and asked to select from a series of response choices that describe how a student would behave in situation. These response choices could describe a range of student responses that do not prioritize the desirability of one behavior over another. For example, a scenario may describe a group science investigation in which one participant does something to destroy the project. Following the scenario, the teacher would be asked to consider prior observations of the student in comparable circumstances, and select among 5 response choices, the response that describes the way the student would most likely behave.

Differences observed in scores based on gender and age at time of rating support the development of norms that account for these factors. Developing norms for students based on racial/ ethnic heritage is not supported given the diversity of student populations and complex ways in which a child's sociocultural experiences influence social and emotional development. Yet, test developers should ensure that the norming group on which standardized rating scores

will be based adequately reflects the racial/ethnic and language diversity within American schools.

**More comprehensive teacher training on assessment practices should be mandated prior to rating students' social and emotional skills and competencies.** If teacher rating scales continue to be pursued as a way of evaluating students social and emotional skills in the school setting, test developers need to do more to ensure that teachers are approaching the cognitive tasks associated with rating students in a more consistent manner. Further instruction is required to *define behavioral terms* on rating scale measures to increase consistency across rater. Test developers may provide examples and non-examples of target behaviors using video vignettes to help calibrate standards of behavior across respondents. Explicit instruction should also be provided to *demonstrate retrieval thought processes* associated with careful and thorough ratings. The application of top-down assessments consistently applied across all behaviors should be discouraged as this type of rating response style may be more susceptible inaccurate ratings caused by 'halo effects' or stereotyped bias. Instead, informants should be encouraged to actively recall specific events or experiences with students to determine ratings. Furthermore, evidence from the present study suggests that teachers would benefit from guidance on how to *interpret frequency ratings* and the importance of *consider ratings across the full range*. Raters should be provided guidance on how to interpret 'frequently' for students in their grade level. There should be agreement on whether frequently means three times per day or three times per week.

### **Future Research**

Additional research is needed to learn more about teacher practices in completing rating scales and child, informant, contextual, and instrument-related factors that are associated with

these practices. Longitudinal studies that track the of ratings of individual teachers over time are required to provide information about teacher response patterns and factors that are associated with these patterns. The current study investigated age, gender, race, intellectual giftedness, and eligibility for support services as child level factors. Future studies are needed to evaluate the impacts of child poverty and recency of immigration on teacher ratings. Additional informant factors such as teacher stress, years of teaching experience, and beliefs about student development should be also investigated to learn how these factors are associated with response patterns and ratings of groups of students. Additionally, further study into how race, cultural heritage, and religious tradition influence teachers' interpretations of student behavior and ratings are warranted.

Future work is also needed to learn how educators and parents are interpreting ratings completed by other informants. What inferences about a student's skills, or abilities are made when a parent or teacher receives a numerical score to describe a child's social and emotional competency? Does the score recipient question the validity of the score? Is the score interpreted as an objective measure of a student's true skill in the school setting? Additionally, research is needed to determine how are schools using teacher ratings or other measures of student social and emotional skills and competencies to inform instruction for individual students or groups of students.

Finally, the academic literature on the assessment of social and emotional skills and competencies is primarily comprised of quantitative studies that evaluate relationships among various factors associated with specific measures of social and emotional skills or competencies. Qualitative studies, in the form ethnographic studies and teacher interviews, are needed to explore how teachers are understanding and responding to student behaviors. There is a sense of

urgency to begin these studies as many teachers are engaging in new practices as they embark on teaching social and emotional skills to diverse groups of students. It is likely that their initiation into this work is fertile with thoughts about student behavior. Research projects are needed to explore how teachers' beliefs, thoughts, values inform and shape their instructional and assessment practices regarding social and emotional learning.

## References

- Abikoff, H., Courtney, M., Jr, W. E. P., & Koplewicz, H. S. (1993). Teachers' ratings of disruptive behaviors: The influence of halo effects. *Journal of Abnormal Child Psychology*, *21*(5), 519–533.  
<https://doi.org/10.1007/BF00916317>
- Abikoff, H., & Gittelman, R. (1985). The normalizing effects of methylphenidate on the classroom behavior of ADDH children. *Journal of Abnormal Child Psychology*, *13*(1), 33–44.  
<https://doi.org/10.1007/BF00918370>
- Academic, Social, and Emotional Learning Act of 2013 (2013 - H.R. 1875). (n.d.). Retrieved July 21, 2016, from <https://www.govtrack.us/congress/bills/113/hr1875>
- Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, *101*(2), 213–232. <https://doi.org/10.1037/0033-2909.101.2.213>
- Advanced Research Seminar on Cognitive Aspects of Survey Methodology, Jabine, T. B., National Research Council (U.S.), & Committee on National Statistics. (1984). *Cognitive aspects of survey methodology building a bridge between disciplines: report of the Advanced Research Seminar on Cognitive Aspects of Survey Methodology*. Washington, D.C.: National Academy Press : Available from the Committee. Retrieved from <http://site.ebrary.com/id/10495478>
- Alexander, K. L., Entwisle, D. R., & Dauber, S. L. (1993). First-Grade Classroom Behavior: Its Short- and Long-Term Consequences for School Performance. *Child Development*, *64*(3), 801–814.  
<https://doi.org/10.2307/1131219>
- Anselm L. Strauss. (1987). *Qualitative analysis for social scientists*. Cambridge Cambridgeshire ; New York: Cambridge University Press.
- Asher, S., & Hymel, S. (1981). Children's social competence in peer relations: Sociometric and behavioral assessment. In *Social Competence* (pp. 125–157). Guilford Press.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *The Psychology of Learning and Motivation*, 89–105.

- Babad, E. Y., Inbar, J., & Rosenthal, R. (1982). Teachers' judgment of students' potential as a function of teachers' susceptibility to biasing information. *Journal of Personality and Social Psychology*, 42(3), 541–547. <https://doi.org/10.1037/0022-3514.42.3.541>
- Baer, P. E., & Bandura, A. (1963). 1. Behavior Theory and Identificatory Learning. *American Journal of Orthopsychiatry*, 33(4), 591–601.
- Bandura, A. (1969). Social learning of moral judgments. *Journal of Personality and Social Psychology*, 11(3), 275.
- Bandura, A. (1977a). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191–215. <https://doi.org/10.1037/0033-295X.84.2.191>
- Bandura, A. (1977b). *Social learning theory*. Englewood Cliffs, N.J.: Prentice Hall.
- Bandura, A. (1978). The self system in reciprocal determinism. *American Psychologist*, 33(4), 344.
- Bandura, A., & McDonald, F. J. (1963). Influence of social reinforcement and the behavior of models in shaping children's moral judgment. *The Journal of Abnormal and Social Psychology*, 67(3), 274–281. <https://doi.org/10.1037/h0044714>
- Bar-On, R., & Parker, J. D. A. (2000). *The handbook of emotional intelligence: theory, development, assessment, and application at home, school, and in the workplace*. San Francisco, Calif.: Jossey-Bass.
- Barry, T. D., Dunlap, S. T., Cotten, S. J., Lochman, J. E., & Wells, K. C. (2005). The Influence of Maternal Stress and Distress on Disruptive Behavior Problems in Boys. *Journal of the American Academy of Child & Adolescent Psychiatry*, 44(3), 265–273. <https://doi.org/10.1097/00004583-200503000-00011>
- Beatty, P. C., & Willis, G. B. (2007). Research Synthesis: The Practice of Cognitive Interviewing. *Public Opinion Quarterly*, 71(2), 287–311. <https://doi.org/10.1093/poq/nfm006>
- Beets, M. W., Flay, B. R., Vuchinich, S., Snyder, F. J., Acock, A., Li, K.-K., ... Durlak, J. (2009). Use of a social and character development program to prevent substance use, violent behaviors, and

- sexual activity among elementary-school students in Hawaii. *American Journal of Public Health*, 99(8), 1438–1445. <https://doi.org/10.2105/AJPH.2008.142919>
- Bellanca, J. A. (2011). *21st Century Skills: Rethinking How Students Learn*. Solution Tree Press.
- Berg-Nielsen, T. S., Solheim, E., Belsky, J., & Wichstrom, L. (2011). Preschoolers' Psychosocial Problems: In the Eyes of the Beholder? Adding Teacher Characteristics as Determinants of Discrepant Parent–Teacher Reports. *Child Psychiatry & Human Development*, 43(3), 393–413. <https://doi.org/10.1007/s10578-011-0271-0>
- Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *American Psychologist*, 57(2), 111–127. <https://doi.org/10.1037/0003-066X.57.2.111>
- Blair, C. (2010). Stress and the Development of Self-Regulation in Context. *Child Development Perspectives*, 4(3), 181–188. <https://doi.org/10.1111/j.1750-8606.2010.00145.x>
- Blankstein, K. R., Toner, B. B., & Flett, G. L. (1989). Test anxiety and the contents of consciousness: Thought-listing and endorsement measures. *Journal of Research in Personality*, 23(3), 269–286. [https://doi.org/10.1016/0092-6566\(89\)90001-9](https://doi.org/10.1016/0092-6566(89)90001-9)
- Block, J. H., & Block, J. (n.d.). The role of ego-control and ego-resiliency in the organization of behavior. In *The Minnesota Symposia on Child Psychology* (Vol. 13, pp. 39–101). Hillsdale, NJ: Erlbaum.
- Bloom, H. S., & Michalopoulos, C. (2011). When is the Story in the Subgroups? *Prevention Science*, 14(2), 179–188. <https://doi.org/10.1007/s11121-010-0198-x>
- Blunden, D., Spring, C., & Greenberg, L. M. (1974). Validation of the Classroom Behavior Inventory. *Journal of Consulting and Clinical Psychology*, 42(1), 84–88. <https://doi.org/10.1037/h0035905>
- Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, 11(2), 177–220. [https://doi.org/10.1016/0010-0285\(79\)90009-4](https://doi.org/10.1016/0010-0285(79)90009-4)
- Brackett, M. A., Mayer, J. D., & Warner, R. M. (2004). Emotional intelligence and its relation to everyday behaviour. *Personality and Individual Differences*, 36(6), 1387–1402. [https://doi.org/10.1016/S0191-8869\(03\)00236-8](https://doi.org/10.1016/S0191-8869(03)00236-8)

- Brackett, M. A., Rivers, S. E., Reyes, M. R., & Salovey, P. (2012). Enhancing academic performance and social and emotional competence with the RULER feeling words curriculum. *Learning & Individual Differences, 22*(2), 218–224. <https://doi.org/10.1016/j.lindif.2010.10.002>
- Bridgeland, J., Bruce, M., & Hariharan, A. (2013). The Missing Piece: A National Teacher Survey on How Social and Emotional Learning Can Empower Children and Transform Schools. A Report for CASEL. Executive Summary.
- Broadbent, D. E., & Broadbent, M. H. P. (1981). Recency effects in visual memory. *The Quarterly Journal of Experimental Psychology Section A, 33*(1), 1–15.  
<https://doi.org/10.1080/14640748108400762>
- Bronfenbrenner, U. (1995). The bioecological model from a life course perspective: Reflections of a participant observer. Retrieved from  
<http://psycnet.apa.org/index.cfm?fa=main.doiLanding&uid=1995-98394-017>
- Bronfenbrenner, U. (2005). *Making human beings human: bioecological perspectives on human development*. Thousand Oaks: Sage Publications.
- Bronfenbrenner, U., & Ceci, S. J. (1994). Nature-nuture reconceptualized in developmental perspective: A bioecological model. *Psychological Review, 101*(4), 568.
- Bronfenbrenner, U., & Morris, P. A. (1998). The ecology of developmental processes. Retrieved from  
<http://psycnet.apa.org/psycinfo/2005-01926-019>
- Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition, 5*(1), 73–99.  
[https://doi.org/10.1016/0010-0277\(77\)90018-X](https://doi.org/10.1016/0010-0277(77)90018-X)
- Canivez, G. L., Watkins, M. W., & Schaefer, B. A. (2002). Interrater agreement for discriminant classifications for the adjustment scales for children and adolescents. *Psychology in the Schools, 39*(4), 375–384. <https://doi.org/10.1002/pits.10043>
- Caprara, G., Barbaranelli, C., Pastorelli, C., Bandura, A., & Zimbardo, P. (2000). Prosocial Foundations of Children's Academic Achievement. *Psychological Science, 11*(4), 302–306.

- Carroll, J. J., & Steward, M. S. (1984). The role of cognitive development in children's understandings of their own feelings. *Child Development*, 55(4), 1486–1492. <https://doi.org/10.2307/1130018>
- Carter, A. S., Briggs-Gowan, M. J., & Davis, N. O. (2004). Assessment of young children's social-emotional development and psychopathology: recent advances and recommendations for practice. *Journal of Child Psychology and Psychiatry*, 45(1), 109–134. <https://doi.org/10.1046/j.0021-9630.2003.00316.x>
- CASEL Guide Online. (n.d.). Retrieved October 23, 2015, from <http://www.casel.org/guide/>
- Caspi, A. (1995). Temperamental Origins of Child and Adolescent Behavior Problems: From Age Three to Age Fifteen. *Child Development*, 66(1), 55–68.
- Castro, F. G., Kellison, J. G., Boyd, S. J., & Kopak, A. (2010). A Methodology for Conducting Integrative Mixed Methods Research and Data Analyses. *Journal of Mixed Methods Research*, 4(4), 342–360. <https://doi.org/10.1177/1558689810382916>
- Catalano, R. F., Berglund, M. L., Ryan, J. A. M., Lonczak, H. S., & Hawkins, J. D. (2004). Positive Youth Development in the United States: Research Findings on Evaluations of Positive Youth Development Programs. *The Annals of the American Academy of Political and Social Science*, 591, 98–124.
- Catalano, R. F., Hawkins, J. D., Berglund, M. L., Pollard, J. A., & Arthur, M. W. (2002). Prevention science and positive youth development: competitive or cooperative frameworks? *Journal of Adolescent Health*, 31(6, Supplement), 230–239. [https://doi.org/10.1016/S1054-139X\(02\)00496-2](https://doi.org/10.1016/S1054-139X(02)00496-2)
- Chafouleas, S. M., Volpe, R. J., Gresham, F. M., & Cook, C. R. (2010). School-Based Behavioral Assessment Within Problem-Solving Models: Current Status and Future Directions. *School Psychology Review*, 39(3), 343–349.
- Children's health care spending report: 2007-2010. (2012). Health Care Cost Institute.

- Chi, T. C., & Hinshaw, S. P. (2002). Mother–Child Relationships of Children with ADHD: The Role of Maternal Depressive Symptoms and Depression-Related Distortions. *Journal of Abnormal Child Psychology*, 30(4), 387–400. <https://doi.org/10.1023/A:1015770025043>
- Cicchetti, D., Ganiban, J., & Barnett, D. (1991). Contributions from the study of high-risk populations to understanding the development of emotion regulation. In J. Garber & K. A. Dodge (Eds.), *The development of emotion regulation and dysregulation* (pp. 15–48). New York, NY, US: Cambridge University Press.
- Coie, J. D., Dodge, K. A., & Coppotelli, H. (1982). Dimensions and types of social status: A cross-age perspective. *Developmental Psychology*, 18(4), 557–570.
- Coie, J. D., Watt, N. F., West, S. G., Hawkins, J. D., Asarnow, J. R., Markman, H. J., ... Long, B. (1993). The science of prevention: A conceptual framework and some directions for a national research program. *American Psychologist*, 48(10), 1013–1022. <https://doi.org/10.1037/0003-066X.48.10.1013>
- Collaborative for Academic, Social, and Emotional Learning (CASEL). (2013). *CASEL Guide 2013: Effective Social and Emotional Learning Programs: Preschool and Elementary Edition*. Chicago, Ill: CASEL. Retrieved from [http://books.google.com/books?hl=en&lr=&id=MuDGDHCb\\_iwC&oi=fnd&pg=PR7&dq=%22for+Academic,+Social,+and+Emotional%22+&ots=oFcZj8UEyZ&sig=N3vPqH4a1A81sSsuwLHgm\\_x8iw8](http://books.google.com/books?hl=en&lr=&id=MuDGDHCb_iwC&oi=fnd&pg=PR7&dq=%22for+Academic,+Social,+and+Emotional%22+&ots=oFcZj8UEyZ&sig=N3vPqH4a1A81sSsuwLHgm_x8iw8)
- Committee for Children. (2013). Second Step Program Kindergarten-Grade 5 Review of Research. Committee for Children. Retrieved from [http://www.cfchildren.org/Portals/0/SS\\_K5/K-5\\_DOC/K-5\\_Review\\_Research\\_SS.pdf](http://www.cfchildren.org/Portals/0/SS_K5/K-5_DOC/K-5_Review_Research_SS.pdf)
- Conners, C. K. (1969). A Teacher Rating Scale for Use in Drug Studies with Children. *American Journal of Psychiatry*, 126(6), 884–888. <https://doi.org/10.1176/ajp.126.6.884>
- Cook, C. R., Rasetshwane, K. B., Truelson, E., Grant, S., Dart, E. H., Collins, T. A., & Sprague, J. (2011). Development and validation of the Student Internalizing Behavior Screener: Examination

- of reliability, validity, and classification accuracy. *Assessment for Effective Intervention*, 36(2), 71–79. <https://doi.org/10.1177/1534508410390486>
- Cooper, B. R., Moore, J. E., Powers, C. J., Cleveland, M., & Greenberg, M. T. (2014). Patterns of early reading and social skills associated with academic success in elementary school. *Early Education and Development*, 25(8), 1248–1264. <https://doi.org/10.1080/10409289.2014.932236>
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research*. Los Angeles: SAGE Publications.
- Darling-Hammond, L., & Friedlaender, D. (2008). Creating Excellent and Equitable Schools. *Educational Leadership*, 65(8), 14.
- Denham, S. A. (1986). Social Cognition, Prosocial Behavior, and Emotion in Preschoolers: Contextual Validation. *Child Development*, 57(1), 194–201. <https://doi.org/10.2307/1130651>
- 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.
- Denham, S. A. (2015). Assessment of SEL in Educational Contexts. In *Handbook of Social and Emotional Learning* (pp. 285–300). New York, NY: Guilford Press.
- Denham, S. A., Ji, P., & Hamre, B. K. (2010). Compendium of Preschool through Elementary School Social Emotional Learning and Associated Assessment Measures. Retrieved from <http://static.squarespace.com/static/513f79f9e4b05ce7b70e9673/t/527815bbe4b057bbfd8adb2e/1383601595064/compendium-of-preschool-through-elementary-school-social-emotional-learning-and-associated-assessment-measures.pdf>
- Denham, S. A., Wyatt, T. M., Bassett, H. H., Echeverria, D., & Knox, S. S. (2009). Assessing social-emotional development in children from a longitudinal perspective. *Journal of Epidemiology and Community Health*, 63(Suppl1), I37–I52. <https://doi.org/10.1136/jech.2007.070797>
- Desimone, L. M., & Le Floch, K. C. (2004). Are We Asking the Right Questions? Using Cognitive Interviews to Improve Surveys in Education Research. *Educational Evaluation and Policy Analysis*, 26(1), 1–22. <https://doi.org/10.3102/01623737026001001>

- Dishion, T. J., McCord, J., & Poulin, F. (1999). When interventions harm: Peer groups and problem behavior. *American Psychologist*, *54*(9), 755.
- Dobbs, J., & Arnold, D. H. (2009). Relationship between preschool teachers' reports of children's behavior and their behavior toward those children. *School Psychology Quarterly*, *24*(2), 95–105. <https://doi.org/10.1037/a0016157>
- Domitrovich, C. E., Bradshaw, C. P., Greenberg, M. T., Embry, D., Poduska, J. M., & Jalongo, N. S. (2009). Integrated models of school-based prevention: Logic and theory. *Psychology in the Schools*, n/a–n/a. <https://doi.org/10.1002/pits.20452>
- Domitrovich, C. E., & Greenberg, M. T. (2000). The Study of Implementation: Current Findings From Effective Programs that Prevent Mental Disorders in School-Aged Children. *Journal of Educational and Psychological Consultation*, *11*(2), 193–221. [https://doi.org/10.1207/S1532768XJEPC1102\\_04](https://doi.org/10.1207/S1532768XJEPC1102_04)
- Downey, D. B., & Pribesh, S. (2004). When Race Matters: Teachers' Evaluations of Students' Classroom Behavior. *Sociology of Education*, *77*(4), 267–282. <https://doi.org/10.1177/003804070407700401>
- 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, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ... Japel, C. (2007). School Readiness and Later Achievement. *Developmental Psychology*, *43*(6), 1428–1446.
- Dunesbury, L., Weissberg, R., & Domitrovich, C. (2014). State Standards to Advance Social and Emotional Learning: Findings from CASEL's State Scan of Social and Emotional Learning Standards, Preschool Through High School, 2014. CASEL.
- 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>

- Durlak, J. A., Weissberg, R. P., & Pachan, M. (2010). A Meta-Analysis of After-School Programs That Seek to Promote Personal and Social Skills in Children and Adolescents. *American Journal of Community Psychology, 45*(3-4), 294–309. <https://doi.org/10.1007/s10464-010-9300-6>
- Dweck, C. S. (2006). *Mindset: the new psychology of success*. New York: Random House.
- Edebrock, C. (1984). Problems and Issues in Using Rating Scales to Assess Child Personality and Psychopathology. *School Psychology Review, 12*, proach.
- Ehrenberg, R. G., Goldhaber, D. D., & Brewer, D. J. (1995). Do Teachers' Race, Gender, and Ethnicity Matter? Evidence from the National Educational Longitudinal Study of 1988. *Industrial and Labor Relations Review, 48*(3), 547–561. <https://doi.org/10.2307/2524781>
- Eisenberg, N., Cumberland, A., Spinrad, T. L., Fabes, R. A., Shepard, S. A., Reiser, M., ... Guthrie, I. K. (2001). The Relations of Regulation and Emotionality to Children's Externalizing and Internalizing Problem Behavior. *Child Development, 72*(4), 1112–1134. <https://doi.org/10.1111/1467-8624.00337>
- Eisenberg, N., Fabes, R. A., Guthrie, I. K., & Reiser, M. (2000). Dispositional emotionality and regulation: Their role in predicting quality of social functioning. *Journal of Personality and Social Psychology, 78*(1), 136–157. <https://doi.org/10.1037/0022-3514.78.1.136>
- Eisenberg, N., Fabes, R. A., Murphy, B., Maszk, P., Smith, M., & Karbon, M. (1995). The Role of Emotionality and Regulation in Children's Social Functioning: A Longitudinal Study. *Child Development, 66*(5), 1360–1384. <https://doi.org/10.2307/1131652>
- Eisenberg, N., Sadovsky, A., & Spinrad, T. L. (2005). Associations of emotion-related regulation with language skills, emotion knowledge, and academic outcomes. *New Directions for Child and Adolescent Development, 2005*(109), 109–118. <https://doi.org/10.1002/cd.143>
- Eisenberg, N., Shepard, S. A., Fabes, R. A., Murphy, B. C., & Guthrie, I. K. (1998). Shyness and Children's Emotionality, Regulation, and Coping: Contemporaneous, Longitudinal, and Across-Context Relations. *Child Development, 69*(3), 767–790. <https://doi.org/10.1111/j.1467-8624.1998.tb06242.x>

- Eisenberg, N., & Spinrad, T. L. (2004). Emotion-Related Regulation: Sharpening the Definition. *Child Development, 75*(2), 334–339.
- Elias, M. J. (1997). *Promoting Social and Emotional Learning: Guidelines for Educators*. ASCD.
- Elias, M. J., Ferrito, J., & Mocerri, D. (2016). *The other side of the report card: assessing students' social, emotional, and character development*. Thousand Oaks, California: Corwin, a SAGE Company.
- Elias, M., Leverett, L, Duffell, J. C., Humphrey, N., Stepney, C., & Ferrito, J. (2015). Integrating SEL with Related Prevention and Youth Development Approaches. In *Handbook of Social and Emotional Learning* (1st ed., pp. 33–49). New York; London: Guilford Press.
- Elliot, S. N., & Gresham, F. M. (1987). Children's Social Skills Assessment and Classification Practises.pdf. *Journal of Counseling and Development, 66*, 96–99.
- Elliott, S. N., Malecki, C. K., & Demaray, M. K. (2001). New directions in social skills assessment and intervention for elementary and middle school students. *Exceptionality, 9*(1-2), 19–32.
- Enterprises, C., Bridgeland, J., Bruce, M., & Hariharan, A. (2013). *MiSSiNg PiECE*. Civic Enterprises, CASEL. Retrieved from <http://s3.documentcloud.org/documents/700800/casel-report.pdf>
- Epstein, J. N., March, J. S., Connors, C. K., & Jackson, D. L. (1998). Racial Differences on the Conners Teacher Rating Scale. *Journal of Abnormal Child Psychology, 26*(2), 109–118.  
<https://doi.org/10.1023/A:1022617821422>
- E. Tory (Edward Tory) Higgins, & Arie W. Kruglanski. (1996). *Social psychology: handbook of basic principles*. New York: Guilford Press.
- Eysenck, M. W. (2013). Anxiety and cognitive performance.
- Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion, 7*(2), 336–353. <https://doi.org/10.1037/1528-3542.7.2.336>
- Fang, Z. (1996). A review of research on teacher beliefs and practices. *Educational Research, 38*(1), 47–65. <https://doi.org/10.1080/0013188960380104>
- Federal Policy and Legislation. (n.d.). Retrieved March 19, 2015, from <http://www.casel.org/federal-policy-and-legislation/>

- Frey, K. S., Nolen, S. B., Van Schoiack Edstrom, L., & 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.  
<https://doi.org/10.1016/j.appdev.2004.12.002>
- Friedlaender, D., Burns, D., Lewis-Charp, H., Cook-Harvey, C. M., & Darling-Hammond, L. (2014, June). Student-Centered Schools: Closing the Opportunity Gap.
- Garner, P., Mahatmya, D., Brown, E., & Vesely, C. (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.  
<https://doi.org/10.1007/s10648-014-9253-7>
- Garner, P. W. (2010). Emotional Competence and its Influences on Teaching and Learning. *Educational Psychology Review, 22*(3), 297–321.
- Ginsburg, H. (1997). *Entering the Child's Mind: The Clinical Interview In Psychological Research and Practice*. Cambridge University Press.
- Ginsburg, HP. (1992). Empirical-Analysis of the cognitive clinical interview, 27(3-4).
- Goleman, D. (1995). *Emotional intelligence*. New York: Bantam Books.
- Graves, S. L., Blake, J., & Kim, E. S. (2012). Differences in Parent and Teacher Ratings of Preschool Problem Behavior in a National Sample The Significance of Gender and SES. *Journal of Early Intervention, 34*(3), 151–165. <https://doi.org/10.1177/1053815112461833>
- Greenberg, M. T. (2010). School-based prevention: current status and future challenges. *Effective Education, 2*(1), 27–52. <https://doi.org/10.1080/19415531003616862>
- Greenberg, M. T., Domitrovich, C., & Bumbarger, B. (2001). The prevention of mental disorders in school-aged children: Current state of the field. *Prevention & Treatment Prevention & Treatment, 4*(1).
- Greenberg, M. T., Domitrovich, C. E., Graczyk, P. A., & Zins, J. E. (2005). The study of implementation in school-based preventive interventions: Theory, research, and practice. *Promotion of Mental*

- Health and Prevention of Mental and Behavioral Disorders 2005 Series V3*. Retrieved from [http://www.researchgate.net/publication/253475340\\_The\\_Study\\_of\\_Implementation\\_in\\_School-Based\\_Preventive\\_Interventions\\_Theory\\_Research\\_and\\_Practice/file/9c96052a8783446310.pdf](http://www.researchgate.net/publication/253475340_The_Study_of_Implementation_in_School-Based_Preventive_Interventions_Theory_Research_and_Practice/file/9c96052a8783446310.pdf)
- Greenberg, M. T., Kusche, C. A., Cook, E. T., & Quamma, J. P. (1995). Promoting emotional competence in school-aged children: The effects of the PATHS curriculum. *Development and Psychopathology*, 7(1), 117.
- Greenwood, C. R. (1991). Longitudinal Analysis of Time, Engagement, and Achievement in At-Risk versus Non-Risk Students. *Exceptional Children*, 57(6), 521–535.
- Gresham, F. M., & Elliot, S. N. (1990). Social Skills Rating System. American Guidance Service.
- Gresham, F. M., & Stuart, D. (1992). Stability of sociometric assessment: Implications for uses as selection and outcome measures in social skills training. *Journal of School Psychology*, 30(3), 223–231. [https://doi.org/10.1016/0022-4405\(92\)90007-R](https://doi.org/10.1016/0022-4405(92)90007-R)
- Gross, D., Fogg, L., Garvey, C., & Julion, W. (2004). Behavior problems in young children: An analysis of cross-informant agreements and disagreements. *Research in Nursing & Health*, 27(6), 413–425. <https://doi.org/10.1002/nur.20040>
- Grusec, J. E. (1992). Social learning theory and developmental psychology: The legacies of Robert Sears and Albert Bandura. *Developmental Psychology*, 28(5), 776–786. <https://doi.org/10.1037/0012-1649.28.5.776>
- Guerra, N. G., & Bradshaw, C. P. (2008). Linking the prevention of problem behaviors and positive youth development: Core competencies for positive youth development and risk prevention. *New Directions for Child and Adolescent Development*, 2008(122), 1–17. <https://doi.org/10.1002/cd.225>
- Haggerty, K., Elgin, J., & Woolley, A. (2011). Social-emotional learning assessment measures for middle school youth. *Social Development Research Group. University of Washington: Raikes Foundation*. Retrieved from <https://audition.prevention.org/Resources/documents/SELTools.pdf>

- Hanish, L. D., Eisenberg, N., Fabes, R. A., Spinrad, T. L., Ryan, P., & Schmidt, S. (2004). The expression and regulation of negative emotions: Risk factors for young children's peer victimization. *Development and Psychopathology*, *16*(2), 335–353.  
<https://doi.org/10.1017/S0954579404044542>
- Hargreaves, A. (2000). Mixed emotions: teachers' perceptions of their interactions with students. *Teaching and Teacher Education*, *16*(8), 811–826. [https://doi.org/10.1016/S0742-051X\(00\)00028-7](https://doi.org/10.1016/S0742-051X(00)00028-7)
- Hastie, R. (1983). Social Inference. *Annual Review of Psychology*, *34*(1), 511.
- Hawkins, D., Catalano, R. F., Kosterman, R., Abbott, R., & Hill, K. (1999). Preventing Adolescent Health Risk Behavior by Strengthening Protection During Childhood. *Archives of Pediatrics & Adolescent Medicine*, *153*(March), 226–234.
- Heathfield, L. T., & Clark, E. (2004). Shifting from categories to services: Comprehensive school-based mental health for children with emotional disturbance and social maladjustment. *Psychology in the Schools*, *41*(8), 911–920. <https://doi.org/10.1002/pits.20047>
- Heckman, J. J., Stixrud, J., & Urzua, S. (2006). *The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior*. National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w12006>
- Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales?: The reference-group effect. *Journal of Personality and Social Psychology*, *82*(6), 903–918. <https://doi.org/10.1037/0022-3514.82.6.903>
- Hinshaw, S. (1992). Externalizing Behavior Problems and Academic Underachievement in Childhood and Adolescence by Hinshaw. *Psychological Bulletin*, *111*(1), 127–155.
- Hinshaw, S. P., Han, S. S., Erhardt, D., & Huber, A. (1992). Internalizing and Externalizing Behavior Problems in Preschool Children: Correspondence Among Parent and Teacher Ratings and Behavior Observations. *Journal of Clinical Child Psychology*, *21*(2), 143.

- Hoge, R. D. (1983). Psychometric Properties of Teacher-Judgment Measures of Pupil Aptitudes, Classroom Behaviors, and Achievement Levels. *Journal of Special Education, 17*(4). Retrieved from <http://offcampus.lib.washington.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=4743772&site=ehost-live>
- Holmes, D. S. (1970). Differential change in affective intensity and the forgetting of unpleasant personal experiences. *Journal of Personality and Social Psychology, 15*(3), 234–239. <https://doi.org/10.1037/h0029394>
- Hoy, W. K., & Fedman, J. A. (1987). Organizational Health: The Concept and Its Measure. *Journal of Research and Development in Education, 20*(4), 30–37.
- Hughes, J. N., Gleason, K. A., & Zhang, D. (2005). Relationship influences on teachers' perceptions of academic competence in academically at-risk minority and majority first grade students. *Journal of School Psychology, 43*(4), 303–320. <https://doi.org/10.1016/j.jsp.2005.07.001>
- Humphrey, N., Curran, A., Morris, E., Farrell, P., & Woods, K. (2007). Emotional Intelligence and Education: A critical review. *Educational Psychology, 27*(2), 235–254. <https://doi.org/10.1080/01443410601066735>
- Individuals with Disabilities Education Act (IDEA). (n.d.). Retrieved March 17, 2015, from <http://www.apa.org/about/gr/issues/disability/idea.aspx>
- Jacob Cohen. (1987). *Statistical power analysis for the behavioral sciences* (Rev. ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Jensen, E. (2009). *Teaching with Poverty in Mind What Being Poor Does to Kids' Brains and What Schools Can Do About It*. Alexandria: ASCD.
- Jersild, A. (1931). Memory for the pleasant as compared with the unpleasant. *Journal of Experimental Psychology, 14*(3), 284–288. <https://doi.org/10.1037/h0074453>

- Jones, S. M., & Bouffard, S. M. (2012). Social and Emotional Learning in Schools: From Programs to Strategies. Social Policy Report. Volume 26, Number 4. *Society for Research in Child Development*. Retrieved from <http://eric.ed.gov/?id=ED540203>
- Jones, S. M., Brown, J. L., & Lawrence Aber, J. (2011). Two-Year Impacts of a Universal School-Based Social-Emotional and Literacy Intervention: An Experiment in Translational Developmental Research: Experiment in Translational Developmental Research. *Child Development*, 82(2), 533–554. <https://doi.org/10.1111/j.1467-8624.2010.01560.x>
- Joseph A. Durlak. (2015). *Handbook of Social and Emotional Learning Research and Practice*. New York: Guilford Publications.
- Juliet M. Corbin author. (2015). *Basics of qualitative research : techniques and procedures for developing grounded theory* (Fourth edition.). Los Angeles: SAGE.
- Kansas Social, Emotional, and Character Development Model Standards. (2012). Kansas State Board of Education.
- Kautz, T., Heckman, J., Diris, R., Weel, B., & Borghans, L. (2014). *Fostering and Measuring Skills Improving Cognitive and Non-cognitive Skills to Promote Lifetime Success*. Paris: OECD Publishing. Retrieved from <http://dx.doi.org/10.1787/5jxsr7vr78f7-en>
- Keefer, K. V. (2015). Self-Report Assessments of Emotional Competencies A Critical Look at Methods and Meanings. *Journal of Psychoeducational Assessment*, 33(1), 3–23. <https://doi.org/10.1177/0734282914550381>
- Kellam, S. G., Ling, X., Merisca, R., Brown, C. H., & Ialongo, N. (1998). The effect of the level of aggression in the first grade classroom on the course and malleability of aggressive behavior into middle school. *Development and Psychopathology*, null(02), 165–185. <https://doi.org/null>
- Kendziora, K., Weissberg, R., & Dunesbury, L. (2011). Strategies for Social and Emotional Learning: Preschool and Elementary Grade Student Learning Standards and Assessment. National Center for Mental Health Promotion and Youth Violence Prevention, Education Development Center, Inc.

- Keogh, B. K., & Burstein, N. D. (1988). Relationship of temperament to preschoolers' interactions with peers and teachers. (includes bibliography). *Exceptional Children*, 54(5), 456.
- Kim, S., Nordling, J. K., Yoon, J. E., Boldt, L. J., & Kochanska, G. (2013). Effortful Control in “Hot” and “Cool” Tasks Differentially Predicts Children’s Behavior Problems and Academic Performance. *Journal of Abnormal Child Psychology*, 41(1), 43–56.
- Konold, T. R., Walthall, J. C., & Pianta, R. C. (2004). The Behavior of Child Behavior Ratings: Measurement Structure of the Child Behavior Checklist Across Time, Informants, and Child Gender. *Behavioral Disorders*, 29(4), 372–383.
- Kraemer, H. C., Measelle, J. R., Ablow, J. C., Essex, M. J., Boyce, W. T., & Kupfer, D. J. (2003). A New Approach to Integrating Data From Multiple Informants in Psychiatric Assessment and Research: Mixing and Matching Contexts and Perspectives. *American Journal of Psychiatry*, 160(9), 1566–1577. <https://doi.org/10.1176/appi.ajp.160.9.1566>
- Ladd, G. W., & Birch, S. H. (1997). The Teacher-Child Relationship and Children’s Early School Adjustment. *Journal of School Psychology*, 35(1), 61–79.
- Lau, A. S. (2006). Making the case for selective and directed cultural adaptations of evidence-based treatments: examples from parent training. *Clinical Psychology: Science and Practice*, 13(4), 295–310.
- LeBuffe, P., & Naglieri, J. (2008). Devereux Student Strengths Assessment (DESSA). Kaplan Early Learning.
- Leibbrand, J. A., & Watson, B. H. (2010). The Road Less Traveled: How the Developmental Sciences Can Prepare Educators to Improve Student Achievement--Policy Recommendations.
- Levitt, J. M., Saka, N., Hunter Romanelli, L., & Hoagwood, K. (2007). Early identification of mental health problems in schools: The status of instrumentation. *Journal of School Psychology*, 45(2), 163–191. <https://doi.org/10.1016/j.jsp.2006.11.005>

- Lipton, M., & Nowicki, S. (2009). The social emotional learning framework (SELF): A guide for understanding brain-based social emotional learning impairments. *The Journal of Developmental Processes*, 4(2), 99–115.
- Li, Y., Marion, S., Perie, M, & Gong, B. (2010). An Approach for Evaluating the Technical Quality of Interim Assessments. *Peabody Journal of Education*, 85, 163–185.
- Lockwood, A. L. (1993, November). A letter to character educators. *Educational Leadership*, p. 72.
- Lopes, P. N., Salovey, P., & Straus, R. (2003). Emotional intelligence, personality, and the perceived quality of social relationships. *Personality and Individual Differences*, 35(3), 641–658.  
[https://doi.org/10.1016/S0191-8869\(02\)00242-8](https://doi.org/10.1016/S0191-8869(02)00242-8)
- MacIntyre, P. D., & Gardner, R. C. (1991). Methods and Results in the Study of Anxiety and Language Learning: A Review of the Literature\*. *Language Learning*, 41(1), 85–117.  
<https://doi.org/10.1111/j.1467-1770.1991.tb00677.x>
- Malecki, C. K., & Elliot, S. N. (2002). Children’s social behaviors as predictors of academic achievement: A longitudinal analysis. *School Psychology Quarterly*, 17(1), 1.
- Marcin Sklad, DIEKSTRA, R., Monique De Ritter, Jehonathan Ben, & Carolien Gravesteyjn. (n.d.). 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. <https://doi.org/10.1002/pits.21641>
- Martin, R. P., Diane, K., Gaddis, L. R., & Moseley, M. (1988). Prediction of elementary school achievement from preschool temperament: Three studies. *School Psychology Review*, 17(1), 125–137.
- Mashburn, A. J., Hamre, B. K., Downer, J. T., & Pianta, R. C. (2006). Teacher and Classroom Characteristics Associated with Teachers’ Ratings of Prekindergartners’ Relationships and Behaviors. *Journal of Psychoeducational Assessment*, 24(4), 367–380.
- Masten, A. S., & Cicchetti, D. (2010). Developmental cascades. *Development and Psychopathology*, 22(03), 491–495. <https://doi.org/10.1017/S0954579410000222>

- McConaughy, S. H. (1993). Evaluating behavioral and emotional disorders with the CBCL, TRF, and YSR cross-informant scales. *Journal of Emotional & Behavioral Disorders, 1*(1), 40.
- McKown, C. (2015). Challenges and Opportunities in the Direct Assessment of Children's Social and Emotional Comprehension. In *Handbook of Social and Emotional Learning* (pp. 320–335). New York, NY: Guilford Press.
- McLeod, J. D., & Nonnemaker, J. M. (2000). Poverty and Child Emotional and Behavioral Problems: Racial/Ethnic Differences in Processes and Effects. *Journal of Health and Social Behavior, 41*(2), 137–161. <https://doi.org/10.2307/2676302>
- Meltzer, H. (1931). The Forgetting of Pleasant and Unpleasant Experiences in Relation to Intelligence and Achievement. *The Journal of Social Psychology, 2*(2), 216–229. <https://doi.org/10.1080/00224545.1931.9918968>
- Mental Health Surveillance Among Children — United States, 2005–2011. (n.d.). Retrieved June 30, 2016, from <http://www.cdc.gov/mmwr/preview/mmwrhtml/su6202a1.htm?viewType=Print&viewClass=Print>
- Merrell, K. W. (2001). Assessment of Children's Social Skills: Recent Developments, Best Practices, and New Directions. *Exceptionality, 9*, 3–2), p.3–18.
- Metcalfe, J., & Mischel, W. (1999). A hot/cool-system analysis of delay of gratification: Dynamics of willpower. *Psychological Review, 106*(1), 3–19. <https://doi.org/10.1037/0033-295X.106.1.3>
- Miller, K., Chepp, V., Willson, S., & Padilla, J. L. (2014). *Cognitive Interviewing Methodology*. John Wiley & Sons.
- Murdock, B. B. J. (1963). Interpolated recall in short-term memory. *Journal of Experimental Psychology, 66*(6), 525–532. <https://doi.org/10.1037/h0046262>
- Murdock, B. B. J. (1999). The Buffer 30 Years Later: Working Memory in a Theory of Distributed Associative Memory (TODAM). In *On human memory: evolution, progress, and reflections on the 30th anniversary of the Atkinson-Shiffrin Model*. New Jersey: Lawrence Erlbaum Associates.

- Naglieri, J., LeBuffe, P., & Shapiro, V. B. (2009). DESSA-Mini Devereux Student Strengths Assessment K-8th Grade. Devereux.
- National Research Council. (2012). *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*. (J. W. Pellegrino, & M. L. Hilton, Eds.). Washington, D.C. : National Academies Press. Retrieved from [http://books.google.com/books?hl=en&lr=&id=k9byCkEGhbcC&oi=fnd&pg=PR1&dq=%22Fifth+Street,+NW,+Keck+360,+Washington,+DC+20001%3B+\(800\)+624-6242+or%22+%22%C2%A9+National+Academy+of+Sciences.+All+rights%22+%22National+Academy+of+Engineering+was+established+in+1964,+under+the%22+&ots=bFCMMFHdwk&sig=8bSUS1n4yTuihNdDSMQrwROAMts](http://books.google.com/books?hl=en&lr=&id=k9byCkEGhbcC&oi=fnd&pg=PR1&dq=%22Fifth+Street,+NW,+Keck+360,+Washington,+DC+20001%3B+(800)+624-6242+or%22+%22%C2%A9+National+Academy+of+Sciences.+All+rights%22+%22National+Academy+of+Engineering+was+established+in+1964,+under+the%22+&ots=bFCMMFHdwk&sig=8bSUS1n4yTuihNdDSMQrwROAMts)
- Newell, A., & Simon, H. A. (1961). Computer Simulation of Human Thinking. *Science*, *134*(3495), 2011–2017.
- Nickerson, A. B., & Fishman, C. (2009). Convergent and Divergent Validity of the Devereux Student Strengths Assessment. *School Psychology Quarterly*, *24*(1), 48–59.
- Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, *35*(4), 250–256. <https://doi.org/10.1037/0022-3514.35.4.250>
- No Child Left Behind Act of 2001, Pub. L. No. 107-110,20, § U.S.C. (2002).
- Norman H. Anderson, author. (1981). *Foundations of information integration theory*. New York: Academic Press.
- Norman K. Denzin. (2007). *On understanding emotion*. New Brunswick, NJ: Transaction Publishers.
- Norman M. Bradburn. (2004). *Asking questions: the definitive guide to questionnaire design-- for market research, political polls, and social and health questionnaires* (Revised edition.). San Francisco, CA: Jossey-Bass.
- Oberle, E., Schonert-Reichl, K. A., Hertzman, C., & Zumbo, B. D. (2014). Social–emotional competencies make the grade: Predicting academic success in early adolescence. *Journal of*

- Applied Developmental Psychology*, 35(3), 138–147.  
<https://doi.org/10.1016/j.appdev.2014.02.004>
- Ontario Symposium on Personality and Social Psychology. (1981). *Social cognition*. Hillsdale, NJ: Erlbaum Associates.
- Pas, E., & Bradshaw, C. (2014). What Affects Teacher Ratings of Student Behaviors? The Potential Influence of Teachers' Perceptions of the School Environment and Experiences. *Prevention Science*, 15(6), 940–950. <https://doi.org/10.1007/s11121-013-0432-4>
- Payton, A. R. (2009). “Mental health, mental illness, and psychological distress: Same continuum or distinct phenomena:” Erratum. *Journal of Health and Social Behavior*, 50(3), 376–376.  
<https://doi.org/10.1177/002214650905000309>
- Payton, J. W., Wardlaw, D. M., Graczyk, P. A., Bloodworth, M. R., Tompsett, C. J., & Weissberg, R. P. (2000). Social and Emotional Learning: A Framework for Promoting Mental Health and Reducing Risk Behavior in Children and Youth. *Journal of School Health*, 70(5), 179–85.
- Peery, J. C. (1979). Popular, Amiable, Isolated Rejected: A Reconceptualization of Sociometric Status in Preschool Children. *Child Development*, 50(4).
- Peng, K., Nisbett, R. E., & Wong, N. Y. C. (1997). Validity Problems Comparing Values Across Cultures and Possible Solutions. *Psychological Methods*, 2(4), 329–344. <https://doi.org/10.1037/1082-989X.2.4.329>
- Perez, J., Petrides, K., & Furnham, A. (2005). Measuring Trait Emotional Intelligence. In *Emotional Intelligence: An International Handbook* (pp. 181–202). Cambridge, MA US: Hogrefe Publishing.
- Petersen, W. (1966). Success Story, Japanese-American Style. *New York Times Magazine*. New York, NY.
- Pigott, R. L., & Cowen, E. L. (2000). Teacher race, child race, racial congruence, and teacher ratings of children's school adjustment. *Journal of School Psychology*, 38(2), 177–196.  
[https://doi.org/10.1016/S0022-4405\(99\)00041-2](https://doi.org/10.1016/S0022-4405(99)00041-2)

- Powell, R., & Kamphaus, R. (1994, March). *The Utility of Parent, Teacher, and Child Ratings for Predicting Externalizing Behavior Problems*. Presented at the National Association of School Psychologists, Seattle WA.
- President's New Freedom Commission on Mental Health, Pub. L. No. 13262 (2002).
- Reitman, W., R. (1964). Information-Processing Models in Psychology. *Science*, *144*(3623), 1192–1198.  
<https://doi.org/http://www.jstor.org/stable/1713024>
- Renk, K., & Phares, V. (2004). Cross-informant ratings of social competence in children and adolescents. *Clinical Psychology Review*, *24*(2), 239–254. <https://doi.org/10.1016/j.cpr.2004.01.004>
- Rice, J. A., Levine, L. J., & Pizarro, D. A. (2007). “Just stop thinking about it”: Effects of emotional disengagement on children’s memory for educational material. *Emotion*, *7*(4), 812–823.  
<https://doi.org/10.1037/1528-3542.7.4.812>
- Rimm-Kaufman, S. E., Larsen, R. A. A., Baroody, A. E., Curby, T. W., Ko, M., Thomas, J. B., ... DeCoster, J. (2014). Efficacy of the Responsive Classroom Approach Results From a 3-Year, Longitudinal Randomized Controlled Trial. *American Educational Research Journal*, *51*(3), 567–603. <https://doi.org/10.3102/0002831214523821>
- Ringwalt, S. (2008). Developmental Screening and Assessment Instruments with an Emphasis on Social and Emotional Development for Young Children Ages Birth through Five. *National Early Childhood Technical Assistance Center (NECTAC)*.
- Rivers, S. E., Brackett, M. A., Reyes, M. R., Elbertson, N. A., & Salovey, P. (2012). Improving the Social and Emotional Climate of Classrooms: A Clustered Randomized Controlled Trial Testing the RULER Approach. *Prevention Science*, *14*(1), 77–87. <https://doi.org/10.1007/s11121-012-0305-2>
- Rocque, M. (2010). Office Discipline and Student Behavior: Does Race Matter? *American Journal of Education*, *116*(4), 557–581. <https://doi.org/10.1086/653629>
- Ruffalo, S. L., & Elliott, S. N. (1997). Teachers’ and Parents’ Rating of Children’s Social Skills: A Closer Look at Cross-Informant Agreements through an Item Analysis Protocol. *School Psychology Review*, *26*(3), 489–501.

- Saft, E. W., & Pianta, R. C. (2001). Teachers' perceptions of their relationships with students: Effects of child age, gender, and ethnicity of teachers and children. *School Psychology Quarterly*, *16*(2), 125.
- Saldana, J. (2015). *The Coding Manual for Qualitative Researchers*. SAGE.
- Salovey, P., & Mayer, J. D. (2004). Emotional intelligence. In P. Salovey, M. A. Brackett, J. D. Mayer, P. (Ed) Salovey, M. A. (Ed) Brackett, & J. D. (Ed) Mayer (Eds.), *Emotional intelligence: Key readings on the Mayer and Salovey model*. (pp. 1–27). Port Chester, NY, US: Dude Publishing.
- Salovey, P., & Sluyter, D. J. (1997). *Emotional Development and Emotional Intelligence: Educational Implications*. Basic Books.
- Schonert-Reichl, K. A., Hanson-Peterson, J., & Hymel, S. (2015). Social and Emotional Learning and Pre-service Teacher Education. In *Handbook of Social and Emotional Learning* (pp. 406–421). New York, NY: Guilford Press.
- Schonert-Reichl, K. A., Smith, V., Zaidman-Zait, A., & Hertzman, C. (2012). Promoting children's prosocial behaviors in school: Impact of the "Roots of Empathy" program on the social and emotional competence of school-aged children. *School Mental Health*, *4*(1), 1–21.
- Schonfeld, D. J., Adams, R. E., Fredstrom, B. K., Weissberg, R. P., Gilman, R., Voyce, C., ... Speese-Linehan, D. (2015). Cluster-randomized trial demonstrating impact on academic achievement of elementary social-emotional learning. *School Psychology Quarterly*, *30*(3), 406–420.  
<https://doi.org/10.1037/spq0000099>
- Schutz, P. A., & Davis, H. A. (2000). Emotions and Self-Regulation During Test Taking. *Educational Psychologist*, *35*(4), 243–256. [https://doi.org/10.1207/S15326985EP3504\\_03](https://doi.org/10.1207/S15326985EP3504_03)
- Schwarz, N. (2007). Cognitive aspects of survey methodology. *Applied Cognitive Psychology*, *21*(2), 277–287. <https://doi.org/10.1002/acp.1340>
- Sears, R. R. (1951). A theoretical framework for personality and social behavior. *American Psychologist*, *6*(9), 476–482. <https://doi.org/10.1037/h0063671>

- Shapiro, E. S., & Heick, P. F. (2004). School psychologist assessment practices in the evaluation of students referred for social/behavioral/emotional problems. *Psychology in the Schools, 41*(5), 551–561. <https://doi.org/10.1002/pits.10176>
- Sheridan, S. M., Maughan, D. P., & Hungelmann, A. (1999). A contextualized framework for social skills assessment, intervention, and generalization. *School Psychology Review, 28*(1), 84.
- Siegel, L., Dragovich, S., & Marholin, D. (1976). The effects of biasing information on behavioral observations and rating scales. *Journal of Abnormal Child Psychology, 4*(3), 221–233. <https://doi.org/10.1007/BF00917760>
- Sosna, T., & Mastergeorge, A. (2005). Compendium of screening tools for early childhood social-emotional development. *Sacramento, CA: California Institute for Mental Health*. Retrieved from [https://www.researchgate.net/profile/Ann\\_Mastergeorge/publication/266276492\\_Compendium\\_of\\_Screening\\_Tools\\_for\\_Early\\_Childhood\\_Social-Emotional\\_Development/links/551bfed50cf2909047b99551.pdf](https://www.researchgate.net/profile/Ann_Mastergeorge/publication/266276492_Compendium_of_Screening_Tools_for_Early_Childhood_Social-Emotional_Development/links/551bfed50cf2909047b99551.pdf)
- Spivack, G., Platt, J. J., & Shure, M. B. (1976). *The problem-solving approach to adjustment*. San Francisco: Jossey-Bass Publishers.
- Stephan, S., Sugai, G., Lever, N., & Connors, E. (2015). Strategies for Integrating Mental Health into Schools via a Multitiered System of Support. *Child And Adolescent Psychiatric Clinics Of North America, 24*(2), 211–+. <https://doi.org/10.1016/j.chc.2014.12.002>
- Stephen W. Raudenbush. (2002). *Hierarchical linear models: applications and data analysis methods* (2nd ed.). Thousand Oaks: Sage Publications.
- Stinnett, T. A., Havey, J. M., & Oehler-Stinnett, J. (1994). Current Test Usage by Practicing School Psychologists: A National Survey. *Journal of Psychoeducational Assessment, 12*(4), 331–350. <https://doi.org/10.1177/073428299401200403>
- Stone, S. L., Speltz, M. L., Collett, B., & Werler, M. M. (2013). Socioeconomic Factors in Relation to Discrepancy in Parent versus Teacher Ratings of Child Behavior. *Journal of Psychopathology and Behavioral Assessment, 35*(3), 314–320. <https://doi.org/10.1007/s10862-013-9348-3>

- Strive Task Force. (2013, August). Beyond Content: Incorporating Social and Emotional Learning into the Strive Framework.
- Sudman, S. (1982). *Asking questions* (1st ed.). San Francisco: Jossey-Bass.
- Supporting Emotional Learning Act (2014 - H.R. 4509). (n.d.). Retrieved July 21, 2016, from <https://www.govtrack.us/congress/bills/113/hr4509>
- Supporting Social and Emotional Learning Act, 497 H.R. § (2015).
- Tabachnick, B. G., & Fidell, L. (2012). *Using Multivariate Statistics* (6th ed.).
- Tesser, A., & Schwarz, N. (2008). *Blackwell Handbook of Social Psychology: Intraindividual Processes*. John Wiley & Sons.
- The Every Student Succeeds Act: Explained. (2015). *Education Week*, 35(14), 17.
- Thomas, D. E., Bierman, K. L., & Powers, C. J. (2011). The influence of classroom aggression and classroom climate on aggressive-disruptive behavior. *Child Development*, 82(3), 751–7. <https://doi.org/10.1111/j.1467-8624.2011.01586.x>
- Thomas, L. B., Shapiro, E. S., DuPaul, G. J., Lutz, J. G., & Kern, L. (2011). Predictors of Social Skills for Preschool Children at Risk for ADHD: The Relationship Between Direct and Indirect Measurements. *Journal of Psychoeducational Assessment*, 29(2), 114–124. <https://doi.org/10.1177/0734282910378478>
- Thomas-Presswood, T. (2008). *Meeting the needs of students and families from poverty: a handbook for school and mental health professionals*. Baltimore: Paul HBrookes PubCo.
- Thorndike, E. I. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, 4(1), 25–29. <https://doi.org/10.1037/h0071663>
- Tory Higgins, W. S., Rholes, C. R., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology*, 13(2), 141–154. [https://doi.org/10.1016/S0022-1031\(77\)80007-3](https://doi.org/10.1016/S0022-1031(77)80007-3)
- Tourangeau, R. (1984). Cognitive science and survey methods. In *Cognitive aspects of survey design: Building a bridge between disciplines* (pp. 73–100). Washington D.C.: National Academy Press.

- Tourangeau, R. (2000). *The psychology of survey response*. Cambridge, UK; New York: Cambridge University Press.
- Tourangeau, R. (2003). Cognitive Aspects of Survey Measurement and Mismeasurement. *International Journal of Public Opinion Research*, 15(1), 3–7. <https://doi.org/10.1093/ijpor/15.1.3>
- Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin*, 103(3), 299–314. <https://doi.org/10.1037/0033-2909.103.3.299>
- Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124–1131.
- United States, Public Health Service, Office of the Surgeon General, Center for Mental Health Services (U.S.), United States, & Substance Abuse and Mental Health Services Administration. (2001). *Mental health: culture, race, and ethnicity : a supplement to Mental health : a report of the Surgeon General*. Rockville, MD: Dept. of Health and Human Services, U.S. Public Health Service.
- Van Horn, M. L., Atkins-Burnett, S., Karlin, E., Ramey, S. L., & Snyder, S. (2007). Parent Ratings of Children's Social Skills: Longitudinal Psychometric Analyses of the Social Skills Rating System. *School Psychology Quarterly*, 22(2), 162–199.
- Van Rooy, D. L., Viswesvaran, C., & Pluta, P. (2005). An Evaluation of Construct Validity: What Is This Thing Called Emotional Intelligence? *Human Performance*, 18(4), 445–462.
- Vygotskii, L. S., & Cole, M. (1978). *Mind in society : the development of higher psychological processes*. Cambridge: Harvard University Press.
- Walker, H. M., & Horner, R. H. (1996). Integrated approaches to preventing antisocial behavior patterns among school-age children and... *Journal of Emotional & Behavioral Disorders*, 4(4), 194.
- Watson, D. L., & Emery, C. (2010). From Rhetoric to Reality: The Problematic Nature and Assessment of Children and Young People's Social and Emotional Learning. *British Educational Research Journal*, 36(5), 767–786.

- Weare, K., & Gray, G. G. (2003). What works in Developing Children's Emotional and Social Competence and Wellbeing. Department for Education and Skills.
- Weare, K., & Nind, M. (2011). Mental health promotion and problem prevention in schools: what does the evidence say? *Health Promotion International*, *26*(suppl 1), i29–i69.  
<https://doi.org/10.1093/heapro/dar075>
- Weissberg, R. P., & Cascarino, J. (2014). Academic Learning + Social-Emotional Learning = National Priority. *Phi Delta Kappan*, *95*(2), 8–13.
- Weissberg, R. P., Gesten, E. L., Rapkin, B. D., Cowen, E. L., Davidson, E., Flores de Apodaca, R., & McKim, B. J. (1981). Evaluation of a social-problem-solving training program for suburban and inner-city third-grade children. *Journal of Consulting and Clinical Psychology*, *49*(2), 251–261.
- Wentzel, K. R. (1993). Does being good make the grade? Social behavior and academic competence in middle school. *Journal of Educational Psychology*, *85*(2), 357.
- Wigelsworth, M., Humphrey, N., Kalambouka, A., & Lendrum, A. (2010). A Review of Key Issues in the Measurement of Children's Social and Emotional Skills. *Educational Psychology in Practice*, *26*(2), 173–186.
- Wilhelm, A. G., & Andrews-Larson, C. (2016). Why Dont Teachers Understand Our Questions? Reconceptualizing Teachers "Misinterpretation" of Survey Items. *AERA Open*, *2*(2), 2/2/2332858416643077. <https://doi.org/10.1177/2332858416643077>
- Willis, G. B. (2005). *Cognitive interviewing : a tool for improving questionnaire design*. Thousand Oaks, Calif: Sage Publications.
- Winsler, A., Kim, Y. K., & Richard, E. R. (2014). Socio-emotional skills, behavior problems, and Spanish competence predict the acquisition of English among English language learners in poverty. *Developmental Psychology*, *50*(9), 2242–2254. <https://doi.org/10.1037/a0037161>
- Wolcott, C. S., & Williford, A. P. (2015). Teacher and TA Ratings of Preschoolers' Externalizing Behavior Agreement and Associations With Observed Classroom Behavior. *Topics in Early Childhood Special Education*, *34*(4), 211–222. <https://doi.org/10.1177/0271121414546008>

- Zaslow, M., Mackintosh, B., Mancoll, S., & Mandell, S. (2015). Federal Policy Initiatives and Children's SEL. In *Handbook of Social and Emotional Learning* (pp. 549–565). New York, NY: Guilford Press.
- Zimmerman, R. S., Khoury, E. L., Vega, W. A., Gil, A. G., & Warheit, G. J. (1995). Teacher and parent perceptions of behavior problems among a sample of African American, Hispanic, and non-Hispanic white students. *American Journal of Community Psychology, 23*(2), 181–197.  
<https://doi.org/10.1007/BF02506935>
- Zins, J., Bloodworth, M., Weissberg, R., & Walberg, H. (2007). The Scientific Base Linking Social and Emotional Learning to School Success. *Journal of Educational and Psychological Consultation, 17*(2-3), 191–210. <https://doi.org/10.1080/10474410701413145>
- Zins, J. E., Bloodworth, M. R., Weissberg, R., & Walberg, H. J. (2004). The Scientific Base linking Social and Emotional Learning to School Success. In *Building academic success on social and emotional learning: what does the research say?* (pp. 3–22). New York: Teachers College Press.

## Appendix A

### Teacher Semi-Structured Interview Protocol

#### **Introduction**

Thank you for agreeing to participate in today's interview. The purpose of the interview is to better understand how teachers form judgments about their students' social and emotional strengths and challenges. The interview today will have three parts. First, I will ask you to think-aloud as you rate 3 students in your classroom. In the second portion of the interview, I will ask you direct questions about the DESSA-Mini rating scale items. Finally, in the third section of our interview I will ask you questions to learn more about how you informally assess your students' social and emotional strengths and challenges. There are no right or wrong responses to the questions I ask you. You can end the interview at any time.

I would like your permission to audiotape our conversation so that I can be sure to accurately capture your responses. The audiotape will not be shared with district personnel. Do I have your consent to audiotape our interview?

Do you have any questions before we begin?

**Present the participant with a written copy of the DESSA-Mini Rating Scale (Image included below)**

I am providing you with a written copy of the DESSA-Mini Rating Scale items that we will be referencing during our meeting today.

# Appendix A



## Devereux Student Strengths Assessment-mini (DESSA-mini)

FORM 1

Jack A. Naglieri, Paul A. LeBuffe, and Valerie B. Shapiro

Child's Name \_\_\_\_\_ Gender \_\_\_\_\_ DOB \_\_\_\_\_ Grade \_\_\_\_\_  
 Person Completing this Form \_\_\_\_\_ Relationship to Child \_\_\_\_\_  
 Date of Rating \_\_\_\_\_ School/Organization \_\_\_\_\_ Classroom/Program \_\_\_\_\_

This form describes a number of behaviors seen in some children. Read the statements that follow the phrase: *During the past 4 weeks, how often did the child...* and place a check mark in the box underneath the word that tells how often you saw the behavior. Answer each question carefully. There are no right or wrong answers. Please answer every item. If you wish to change your answer, put an X through it and fill in your new choice as shown to the right.

Never	Rarely	Occasionally	Frequently	Very Frequently
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
0	1	2	3	4

Item #	During the past 4 weeks, how often did the child...	Never	Rarely	Occasionally	Frequently	Very Frequently	Score
1.	accept responsibility for what she/he did?	0	1	2	3	4	_____
2.	do something nice for somebody?	0	1	2	3	4	_____
3.	speak about positive things?	0	1	2	3	4	_____
4.	pay attention?	0	1	2	3	4	_____
5.	contribute to group efforts?	0	1	2	3	4	_____
6.	perform the steps of a task in order?	0	1	2	3	4	_____
7.	show care when doing a project or school work?	0	1	2	3	4	_____
8.	follow the advice of a trusted adult?	0	1	2	3	4	_____

Raw Score Sum \_\_\_\_\_  
 Turn over to finish scoring →

Recommendations: \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_

**Apperson** Copyright ©2014, The Institute for Education, All rights reserved.  
**SEL+** No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage or retrieval system, without permission in writing from the publisher.  
Apperson, Inc. 2000 Sowers Creek Road, Charlotte, NC 28210 www.apperson.com (800) 634-6362 Product Code 140000  
 Original DESSA-mini forms are printed in red and black. If this form is only in black and white, it is an unauthorized photocopy, which is a violation of the copyright.

### Think-Aloud Directions

During this first part of the interview, I will be asking you to think-aloud as you go through the process of rating 3 students in your classroom. Please try to say all of your thoughts aloud—just tell me everything that comes to mind, whether it seems important or not (Willis, 2005). I will be taking lots of notes as you talk. Do you have any questions?

Ok let's begin with student # 1. Please silently select a student who was rated in the 'strength' range this fall.

What is the student's gender?

What is the student's race/ ethnicity?

Does the student receive any special services (e.g. SPED, free/reduced lunch, ELL services)?

## Appendix A

OK go ahead and begin think out loud to the extent possible so we know what thoughts come to mind as you rate this student.

**After completing the rating on the first student, ask teacher to select another student in the typical range and repeat the process. The last student selected should be in the ‘need’ range.**

Thank you for sharing your thoughts as you rated students in your classroom. In the next portion of the interview, I will be asking you direct questions about 3 DESSA-Mini items. Please do your best to share your thinking as much as possible, even if you believe that your thoughts are off topic or not important. Do you have any questions before we begin?

### **Verbal-Probing Directions**

I will be asking you a number of direct questions about 3 of the DESSA-Mini items that you just used to rate your students. Please share as much of your thinking as possible.

What does it mean to pay attention?

Please complete the following sentence. Paying attention means \_\_\_\_\_.

Briefly describe at least three instances in which you observed a student paying attention.

Please name three activities in which you are likely observe your students paying attention?

In recalling how frequently a student pays attention, do you think specifically about recent experiences, do you average together experiences you’ve had over a number of weeks, do you think of a particular event, or do you think about it in some other way?

How do you think about the frequency of a student paying attention?

How would you distinguish between a student who pays attention occasionally and one who pays attention frequently?

To what extent do you fully comprehend what paying attention means on this measure?

How confident do you feel in your ability to accurately rate the frequency of a student paying attention?

**The question stems would be repeated for 2 additional items on the DESSA-Mini with the following phrases substituting for the underlined text above:**

**Do something nice for somebody**

**Contribute to group efforts**

**The following prompts can be used to solicit an expanded response or explanation.**

“Tell me more.”

“You said \_\_\_\_\_. Please say more about that.”

### **Expansive Probing Directions**

In addition to this formal assessment of students’ social and emotional strengths, I am interested in learning more about your thoughts pertaining to assessment and the ways in which you

## Appendix A

informally assess students' social and emotional strengths and challenges. I will be asking you open-ended questions to get your thoughts.

What are some essential social and emotional skills that students must develop to be socially and emotionally competent?

How easy or difficult is it for you to evaluate students' social and emotional skills?

The DESSA-Mini asks teachers to base ratings on students' behaviors in the 'past 4 weeks.' How do you think about this timeframe when rating students' behaviors?

How is the assessment of students' social and emotional skills the same as or different from the assessment of students' academic skills?

How can you tell if a student is developing skills—if he or she is progressing?

When rating students within your classroom how does race or cultural heritage inform or not inform your thinking?

What do you see as the pros and cons of evaluating all students social and emotional skills in the school setting?

Do you believe that teachers should be evaluating students' social and emotional skills?

Do you think an evaluation of students' social and emotional skills should be part of their report card and permanent record?

Are rating scales an appropriate way to evaluate students' social and emotional skills?

Thank you for participating in today's interview. Do you have any questions?

**End of interview.**

## Appendix B

### Interview Template

#### Demographic Information

Research ID#:

Site ID#

Gender:

Male

Female

No Response

Race/Ethnicity:

American Indian/ Alaska Native

Asian

Black or African-American

Hawaiian/ Pacific Islander

Hispanic Latino

Multi-Ethnic

White

No Response

Years Teaching:

Prior Schools:

Think-aloud
-------------

#### Student #1 Strength Range

Gender: Male Female No Response

Race/Ethnicity: American Indian/ Alaska Native, Asian, Black or African-American, Hawaiian/ Pacific Islander, Hispanic Latino, Multi-Ethnic, White, No Response

Special Services: SPED, ELL, 504, Free/ Reduced Lunch, Unknown, None

What three words would you use to describe this student?

## Appendix B

### Student #2 Typical Range

Gender: Male Female No Response  
Race/Ethnicity: American Indian/ Alaska Native, Asian, Black of African-American, Hawaiian/ Pacific Islander, Hispanic Latino, Multi-Ethnic, White, No Response  
Special Services: SPED, ELL, 504, Free/ Reduced Lunch, Unknown, None

What three words would you use to describe this student?

### Student #3 Need Range

Gender: Male Female No Response  
Race/Ethnicity: American Indian/ Alaska Native, Asian, Black of African-American, Hawaiian/ Pacific Islander, Hispanic Latino, Multi-Ethnic, White, No Response  
Special Services: SPED, ELL, 504, Free/ Reduced Lunch, Unknown, None

What three words would you use to describe this student?

## Appendix B

## Appendix B

Verbal Probing
----------------

1. What does it mean to pay attention?
  
2. Please complete the following sentence. Paying attention means\_\_\_\_\_.
  
3. Briefly describe at least three instances in which you observed a student paying attention.
  - a.
  
  - b.
  
  - c.
  
4. In what contexts do you observe students paying attention?
  
5. Please name three activities in which you are likely observe your students paying attention?
  - a.
  
  - b.
  
  - c.
  
6. How do you think about the frequency of a student paying attention?

## Appendix B

7. How would you distinguish between a student who pays attention occasionally and one who pays attention frequently?
  
8. To what extent do you fully comprehend what “paying attention” means on this measure?
  1. No comprehension
  2. Uncertain of comprehension
  3. Somewhat comprehend
  4. Comprehend
  5. Fully comprehend
  
9. How confident do you feel in your ability to accurately rate the frequency of a student paying attention?
  1. Not confident
  2. Minimally confident
  3. Somewhat confident
  4. Confident
  5. Fully confident



## Appendix B

7. How would you distinguish between a student who does something nice for somebody occasionally and one who does something nice for somebody frequently?
  
8. To what extent do you fully comprehend what “doing something nice for somebody” means on this measure?
  - a. No comprehension
  - b. Uncertain of comprehension
  - c. Somewhat comprehend
  - d. Comprehend
  - e. Fully comprehend
  
9. How confident do you feel in your ability to accurately rate the frequency of a student doing something nice for somebody?
  - a. Not confident
  - b. Minimally confident
  - c. Somewhat confident
  - d. Confident
  - e. Fully confident

## Appendix B

1. What does it mean to contribute to group efforts?
  
2. Please complete the following sentence. Contributing to group efforts means \_\_\_\_\_.
  
3. Briefly describe at least three instances in which you observed a student contributing to group efforts.
  - a.
  
  - b.
  
  - c.
  
4. In what contexts do you observe students contributing to group efforts?
  
5. Please name three activities in which you are likely observe your students contributing to group efforts?
  - a.
  
  - b.
  
  - c.
  
6. How do you think about the frequency of a student contributing to group efforts?

## Appendix B

7. How would you distinguish between a student who contributes to group efforts occasionally and one who contributes to group efforts frequently?
  
8. To what extent do you fully comprehend what “contributing to group efforts” means on this measure?
  - a. No comprehension
  - b. Uncertain of comprehension
  - c. Somewhat comprehend
  - d. Comprehend
  - e. Fully comprehend
  
9. How confident do you feel in your ability to accurately rate the frequency of a student contributing to group efforts?
  - a. Not confident
  - b. Minimally confident
  - c. Somewhat confident
  - d. Confident
  - e. Fully confident

## Appendix B

Expansive Probing
-------------------

1. What do you consider to be some essential skills students much develop to be socially and emotionally competent?
2. How easy or difficult is it for you to evaluate students' social and emotional skills?
3. The DESSA-Mini asks teachers to base ratings on students' behavior in the "past 4 weeks." How do you think about this timeframe when rating students' behaviors?
4. How is the assessment of students' social and emotional skills the same as or different from the assessment of the of students' academic skills?
5. How can you tell if a student is developing skills—if he or she is progressing?
6. When rating students within your classroom how does race or cultural heritage inform or not inform your thinking?

## Appendix B

7. What do you see as the pros and cons of evaluating all students' social and emotional skills in the school setting?

## Appendix C

### Classroom characteristics for participating teachers' classrooms

**Table 5.24.** *Classroom characteristics of participating teachers*

<i>Int.</i>	<i>%Fem</i>	<i>%Sped</i>	<i>504%</i>	<i>%ELL</i>	<i>%Asian</i>	<i>%Black</i>	<i>%White</i>	<i>%Multi</i>	<i>%Latino</i>
<u>Title</u>									
2	45.0	5.0	0.0	30.0	10.0	10.0	45.0	0.0	35.0
3	38.1	4.8	4.8	33.3	9.5	9.5	28.6	14.3	38.1
4	59.1	4.5	0.0	22.7	9.1	13.6	40.9	4.5	31.8
9	50.0	9.1	4.5	36.4	18.2	13.6	40.9	0.0	27.3
12	50.0	10.0	15.0	25.0	15.0	10.0	40.0	5.0	30.0
14	52.4	14.3	23.8	23.8	23.8	23.8	23.8	23.8	23.8
<u>Non-Title</u>									
1*	52.0	0.0	4.0	0.0	64.0	0.0	32.0	0.0	4.0
5	52.0	4.0	0.0	12.0	36.0	0.0	56.0	0.0	8.0
6	50.0	3.8	0.0	11.5	38.5	0.0	46.2	7.7	3.8
7*	35.0	0.0	0.0	0.0	70.0	0.0	25.0	5.0	0.0
8*	50.0	0.0	0.0	0.0	76.9	3.8	11.5	7.7	0.0
10	68.2	0.0	0.0	18.2	18.2	9.1	50.0	9.1	13.6
11	54.2	8.3	0.0	20.8	37.5	4.2	37.5	8.3	12.5
13	54.2	0.0	4.2	12.5	54.2	4.2	29.2	8.3	4.2
15	60.9	0.0	0.0	13.0	17.4	0.0	60.9	17.4	4.3

\* self-contained gifted classroom

## Appendix D

### Atlas ti Code List and Definitions

#### **\*\*Expansive Probe**

Portion of interview using the expansive probe cognitive interview methodology.

#### **\*\*Think-aloud**

Portion of interview using the think-aloud cognitive interview methodology.

#### **\*Verbal Probe**

Portion of interview using the verbal probe cognitive interview methodology.

#### **\*Def: Attention**

Portion of interview in which teacher is explaining what it means to pay attention.

#### **\*Def: Group**

Portion of interview in which teacher is explaining what it means to contribute to group efforts.

#### **\*Def: Nice**

Portion of interview in which teacher is explaining what it means to do something nice for somebody.

#### **\*EP\_Appropriateness**

Portion of interview in which teacher responds to whether she believes that teachers should be evaluating students' social and emotional skills.

#### **\*EP\_Assessment**

Portion of interview in which teacher explains how the assessment of social and emotional skills is the same or different than the assessment of academic skills.

#### **\*EP\_Developing**

Portion of interview in which teacher explains how she evaluates whether students are developing skills.

#### **\*EP\_EaseEval**

Portion of interview in which teacher explains how easy or difficult it is for her to evaluate r students' social and emotional skills.

#### **\*EP\_Heritage**

Portion of interview in which teacher explains how race or cultural heritage informs her assessment of students' social and emotional skills.

## Appendix D

### **\*EP\_ProCons**

Portion of interview in which teacher describes the pros and cons of evaluating students' social and emotional skills in the school setting.

### **\*EP\_Rating Scale**

Portion of interview in which teacher responds to whether rating scales are an appropriate way to assess students' social and emotional skills.

### **\*EP\_Record**

Portion of interview in which teacher explains whether social and emotional evaluation data should be part of the school record.

### **\*EP\_Timeframe**

Teacher explains how she thinks about the 4-week time-frame that DESSA-Mini utilizes for evaluating students' social and emotional skills.

### **\*Freq: Attention**

Portion of interview in which teacher is explaining how he/she interprets and applies frequency of paying attention.

### **\*Freq: Group**

Portion of interview in which teacher is explaining how he/she interprets and applies frequency of contributing to group efforts.

### **\*Freq: Nice**

Portion of interview in which teacher is explaining how he/she interprets and applies frequency of doing something nice for somebody.

### **\*Need**

Segment of think-aloud in which teacher rates student in need range.

### **\*Strength**

Segment of think-aloud in which teacher rates student in strength range.

### **\*Typical**

Segment of think-aloud in which teacher rates student in typical range.

### **1Need: Attention**

Teacher think-aloud response when rating student in the need range for paying attention.

## Appendix D

### **1Need: Group**

Segment in think-aloud in which teacher rates student in need range for contributing to group efforts.

### **1Need:Nice**

Segment in think-aloud in which teacher rates student in need range on doing something nice for somebody.

### **1Strength:Attention**

Teacher think-aloud response when rating student in the strength range for paying attention.

### **1Strength: Group**

Segment in think-aloud in which teacher rates student in strength range for contributing to group efforts.

### **1Strength:Nice**

Segment in think-aloud in which teacher rates student in strength range on doing something nice for somebody.

### **1Typical:Attention**

Teacher think-aloud response when rating student in the typical range for paying attention.

### **1Typical:Group**

Segment in think-aloud in which teacher rates student in typical range for contributing to group efforts.

### **1Typical:Nice**

Segment in think-aloud in which teacher rates student in typical range on doing something nice for somebody.

### **ADHD**

Student identified as having ADHD.

### **Attention\_Def\_Application of Understanding**

Attention defined as the ability to apply what was taught or experienced to produce verbal or written products. Examples could include participation in classroom discussion, ability to complete written work.

### **Attention\_Def\_Application of Understanding\_NOT**

Recognition that a child may attend to appropriate learning opportunity, but may be unable to understand or demonstrate knowledge through verbal or written output.

## Appendix D

### **Attention\_Def\_Attending to Directions and Conversation**

Attention defined as attending to the teacher giving directions, teaching a lesson, or providing guidance or direction. Attending to conversation could also include attending to a fellow student who is verbally contributing to classroom or group discussion.

### **Attention\_Def\_Awareness**

Awareness describes attention to stimuli, but differs from engagement in that it is more passive and less active. Awareness code also lacks the characteristic of selectively attending to one stimulus over another.

### **Attention\_Def\_Engagement**

Active participation in opportunity for learning. Often describing ability to selectively attend to learning opportunities within the classroom rather than stimuli that distracts from academic learning (e.g. social interactions and side conversations)

### **Attention\_Def\_Physical Behavior**

Attention requires exhibiting specific physical behaviors (e.g. nodding, eyes on speaker, sitting up straight). Several of the behaviors specifically taught through SLANT program.

### **Attention\_Def\_Physical Behavior\_NOT**

Teacher specifies that attention does NOT require students to exhibit specific physical behaviors (e.g. nodding, eyes on speaker, sitting up straight). Students may be coloring, doodling, or even rolling on floor and still be attending

### **Attention\_Def\_Social Emotional Cues**

When defining paying attention, teacher addresses attending to social and emotional cues or the behaviors of others.

### **Comprehension\_\_Uncertain of Comprehension\_Attention**

Teacher rates the extent to which she fully comprehends what paying attention means on this measure as 'Uncertain of Comprehension.'

### **Comprehension\_Comprehend\_Attention**

Teacher rates the extent to which she comprehends what paying attention means on this measure as 'Comprehend.'

### **Comprehension\_Comprehend\_Group**

Teacher rates the extent to which she comprehends what contributing to a group means on this measure as 'Comprehend.'

## Appendix D

### **Comprehension\_Comprehend\_Nice**

Teacher rates the extent to which she comprehends what doing something nice for somebody means on this measure as 'Comprehend.'

### **Comprehension\_Fully Comprehend\_Attention**

Teacher rates the extent to which she fully comprehends what paying attention means on this measure as 'Fully Comprehend.'

### **Comprehension\_Fully Comprehend\_Group**

Teacher rates the extent to which she fully comprehends what contributing to group effort means on this measure as 'Fully Comprehend.'

### **Comprehension\_Fully Comprehend\_Nice**

Teacher rates the extent to which she fully comprehends what doing something nice for somebody means on this measure as 'Fully Comprehend.'

### **Comprehension\_Somewhat Comprehend\_Nice**

Teacher rates the extent to which she fully comprehends what doing something nice for somebody means on this measure as 'Somewhat Comprehend.'

### **Comprehension\_Somewhat Comprehend\_Attention**

Teacher rates the extent to which she fully comprehends what paying attention means on this measure as 'Somewhat Comprehend.'

### **Comprehension\_Somewhat Comprehend\_Group**

Teacher rates the extent to which she fully comprehends what contributing to a group effort means on this measure as 'Somewhat Comprehend.'

### **Confidence\_Confident\_Attention**

Teacher rates her confidence in her ability to accurately rate the frequency of a student paying attention as 'confident.'

### **Confidence\_Confident\_Group**

Teacher rates her confidence in her ability to accurately rate the frequency of a student contributing to a group effort as 'confident.'

### **Confidence\_Confident\_Nice**

Teacher rates her confidence in her ability to accurately rate the frequency of a student doing something nice for somebody as 'confident.'

## Appendix D

### **Confidence\_Fully Confident\_Attention**

Teacher rates her confidence in her ability to accurately rate the frequency of a student paying attention as ‘fully confident.’

### **Confidence\_Fully Confident\_Group**

Teacher rates her confidence in her ability to accurately rate the frequency of a student contributing to a group effort as ‘fully confident.’

### **Confidence\_Fully Confident\_Nice**

Teacher rates her confidence in her ability to accurately rate the frequency of a student doing something nice for somebody as ‘fully confident.’

### **Confidence\_No Confidence\_Attention**

Teacher rates the extent to which she feels confident in her ability to accurately rate the frequency of a student paying attention as ‘not confident’.

### **Confidence\_No Confidence\_Group**

Teacher rates the extent to which she feels confident in her ability to accurately rate the frequency of a student contributing to a group effort as ‘not confident’.

### **Confidence\_No Confidence\_Nice**

Teacher rates the extent to which she feels confident in her ability to accurately rate the frequency of a student doing something nice for somebody as ‘not confident’.

### **Confidence\_Somewhat Confident\_Attention**

Teacher reports he/she is somewhat confident in ability to accurately rate the frequency of student paying attention.

### **Confidence\_Somewhat Confident\_Group**

Teacher reports he/she is somewhat confident in ability to accurately rate the frequency of student contributing to group efforts.

### **Confidence\_Somewhat Confident\_Nice**

Teacher reports he/she is somewhat confident in ability to accurately rate the frequency of student doing something nice for somebody.

### **Descriptor**

Adjectives used to describe individual child rated during think-aloud portion of cognitive interview.

## Appendix D

### **Descriptor\_Ability**

Descriptor references student abilities. Teacher may describe student as bright, capable, intelligent, etc.

### **Descriptor\_Achievement**

Descriptor references traits associated with academic achievement. Teacher may describe student as responsible, efficient, hardworking, motivated, etc. Adjectives associated with perseverance coded separately.

### **Descriptor\_Confident**

Descriptor references student confidence or self-understanding.

### **Descriptor\_Creative**

Descriptor references student creativity or imagination.

### **Descriptor\_Empathy\_Care for Others**

Descriptors references student's care or concern for others. Adjectives may include thoughtful, empathetic, caring, etc.

### **Descriptor\_Flexibility**

Descriptor references student flexibility or ability to adjust to changing circumstances.

### **Descriptor\_Immature Undercontrolled**

Descriptor references student behaviors that are young, immature, impulsive or under controlled.

### **Descriptor\_Need**

Descriptor that teacher provided for student in the need range.

### **Descriptor\_Neg Emot**

Descriptor references student as exhibiting emotional negativity (e.g. frustration, anger, etc.).

### **Descriptor\_Perseverance**

Descriptor references student's strength or persistence. In some instances, the behavior is viewed as an asset and in others, as a hindrance.

### **Descriptor\_Pleas Emot**

Descriptor references student as exhibiting pleasant emotions (e.g. fun, playful, social, positive, charming).

### **Descriptor\_Sensitive Quiet**

Descriptor references student as quiet, sensitive, or timid.

## Appendix D

### **Descriptor\_Strength**

Descriptor that teacher provided for student in the strength range.

### **Descriptor\_Typical**

Descriptor that teacher provided for student in the typical range.

### **Descriptor\_Uncategorized**

Code assigned to descriptors that do not fit into a descriptor category group.

### **Essential Skills**

Skills teachers identified as essential in developing social and emotional competency.

### **Essential\_Decision Making**

Teacher references importance of making conscious decision to behave in a certain way.

### **Essential\_Emotion Expression**

Teacher references importance of student's ability to express emotions in an appropriate manner.

### **Essential\_Emotion Labeling**

Teacher references importance of naming or labeling personal emotions.

### **Essential\_Emotion Recognition**

Teacher references importance of student's ability to recognize personal emotions.

### **Essential\_Emotion Regulation**

Teacher references importance of student's ability to regulate personal emotions.

### **Essential\_Emotion Understanding**

Teacher references importance of understanding personal emotion.

### **Essential\_Empathy**

Description of essential skills includes empathy, kindness, or thoughtfulness toward others.

### **Essential\_Flexibility**

Teacher references importance of being flexible and adaptable to different people or circumstances.

### **Essential\_Perseverance**

Teacher references importance of perseverance, determination or learning from failure.

## Appendix D

### **Essential\_Perspective Taking and Conflict Resolution**

Teacher references student's ability to understand others' perspectives and resolve conflicts as an essential skill.

### **Essential\_Reciprocal Communication**

Teacher references student ability to listen to ideas and express personal ideas with others. Communication surrounding emotions is coded under the Essential\_Emotion Expression code.

### **Essential\_Respect**

Teacher references respect as an essential skill for students.

### **Essential\_Self-Awareness**

Teacher references self-awareness as an essential skill.

### **Essential\_Social Awareness**

Teacher references social awareness as an essential skill.

### **Factor\_Child\_Ability**

Teacher mentions child's capabilities about intellect.

### **Factor\_Child\_Ability\_EP**

Teacher references student academic ability in the expansive probe (EP) section of cognitive interview.

### **Factor\_Child\_Academic Struggle**

Teacher mentions student's academic challenges or lack of understanding about lesson or assignment.

### **Factor\_Child\_Achievement**

Teacher references student's positive academic achievement and/or performance as a student. Achievement is coded distinctly from mention of ability.

### **Factor\_Child\_Achievement\_EP**

Teacher references student academic achievement in the expansive probe (EP) section of cognitive interview.

### **Factor\_Child\_Compliant**

Teacher characterizes student behavior as compliant. The student responds to teacher directions.

## Appendix D

### **Factor\_Child\_Compliant\_EP**

Teacher reverences student compliance in the expansive probe (EP) section of cognitive interview.

### **Factor\_Child\_Compliant\_NOT**

Teacher characterizes student behavior as not compliant. The student does not consistently respond to teacher direction.

### **Factor\_Child\_Cultural Background**

Teacher reverences student's cultural background, race, or ethnicity during think-aloud portion of cognitive interview.

### **Factor\_Child\_Cultural Background\_EP**

Teacher reverences student's cultural background, race, or ethnic heritage during expansive probe portion of cognitive interview.

### **Factor\_Child\_Developmental Level or Age**

Teacher reverences student's age or developmental level during think-aloud portion of cognitive interview.

### **Factor\_Child\_Developmental Level or Age\_EP**

Teacher reverences student's age or developmental level during expansive probe portion of cognitive interview.

### **Factor\_Child\_Effort\_High**

Teacher describes student as consistently putting forth strong effort to succeed in academic work.

### **Factor\_Child\_Effort\_Low**

Teacher describes student as putting forth little effort to succeed in academic work.

### **Factor\_Child\_Effort\_Mixed**

Teacher describes student as putting forth mixed or inconsistent effort to succeed in academic work.

### **Factor\_Child\_Family School Difference**

Teacher describes differences between-school and family expectations for student behavior and/or the way behaviors should be addressed at school.

## Appendix D

### **Factor\_Child\_Family School Difference\_EP**

Teacher references differences between-school and family responses to student behavior during expansive probe portion of cognitive interview.

### **Factor\_Child\_Family Stressors**

Teacher mentions family circumstances that create stress for student. Stressors may include family conflict, divorce, addiction, refugee status, poverty.

### **Factor\_Child\_Family Stressors\_EP**

Teacher references family stressors that impact students during expansive probe portion of cognitive interview.

### **Factor\_Child\_Family Support**

Teacher describes child as having a family that is engaged in supporting student's well-being, academic achievement, or positive relationships at school.

### **Factor\_Child\_Family Support\_EP**

Teacher references family support and its impact on students during expansive probe portion of cognitive interview.

### **Factor\_Child\_Gender**

Teacher mentions child's gender in the context of rating a student's social and emotional skills.

### **Factor\_Child\_Gender\_EP**

Teacher references student gender during expansive probe portion of cognitive interview.

### **Factor\_Child\_Language**

Teacher mentions child's language development as a factor in exhibiting social and emotional skills.

### **Factor\_Child\_StressFatigue**

Teacher references child fatigue or stress in children's ability to perform social and emotional skills.

### **Factor\_Child\_Temperament Positive**

Teacher describing student's nature as having positive emotionality.

### **Factor\_Child\_Temperament\_EP**

Teacher references temperament or personality during the expansive probe portion of cognitive interview.

## Appendix D

### **Factor\_Child\_Temperament\_Extroversion**

Teacher describes student as very social and likely to seek social interactions with others.

### **Factor\_Child\_Temperament\_Negative**

Teacher characterization of student as having negative emotionality or being uncooperative.

### **Factor\_Child\_Temperament\_Quiet**

Teacher describes child as quiet.

### **Factor\_Child\_Temperament\_Quiet\_EP**

Teacher reverences quiet or introverted children during the expansive probe portion of cognitive interview.

### **Factor\_Child\_Temperament\_Withdrawn**

Teacher characterizes child as socially withdrawn.

### **Factor\_Context\_Adult Modeling**

Teacher references adult modeling in the development or display of students' social and emotional skills.

### **Factor\_Context\_Classroom.School Climate**

Teacher references the classroom or school social and emotional climate as influencing the acquisition or display of social and emotional skills.

### **Factor\_Context\_New Student**

Teacher references student's newness to school or program as factor in student behavior.

### **Factor\_Context\_Peer Influence**

Teacher reference the impact of peer interactions on the acquisition or display of social and emotional skills.

### **Factor\_Context\_Setting Demands**

Teacher references setting demands (e.g. such as content area, time of day, large group/small group) on the display of social and emotional skills.

### **Factor\_Context\_Single Context**

Teacher references the limitations of evaluating social and emotional skills based on observations in a single setting or context.

### **Factor\_Context\_Teacher**

Teacher references classroom context, but not about behaviors of individual students.

## Appendix D

### **Factor\_Informant\_Childhood Experience**

Teacher references personal childhood experiences.

### **Factor\_Informant\_Culture**

Teacher references own culture and how it may influence perspective on student behavior.

### **Factor\_Informant\_Disposition**

Teacher references personal disposition, mood or temperament.

### **Factor\_Informant\_EQ Ability**

Teacher references social and emotional abilities of self or other educators.

### **Factor\_Informant\_Not Classified**

Code assigned to informant factors that do not fall into other informant code categories.

### **Factor\_Informant\_Parent Experience**

Teacher references personal experiences as a parent.

### **Factor\_Informant\_Practices**

Teacher references personal classroom practices.

### **Factor\_Informant\_Relationship\_Conflict**

Code references teacher student relationship as conflicted.

### **Factor\_Informant\_Relationship\_Conflict\_EP**

Code references teacher mention of teacher-student conflict during expansive probe section of interview.

### **Factor\_Informant\_Relationship\_Cooperative**

Code references teacher student relationship and characterizes as cooperative or close.

### **Factor\_Informant\_Relationship\_Dependent**

Code references teacher student relationship and characterizes as dependent. The student is excessively needy of teacher attention.

### **Factor\_Informant\_Relationship\_EP**

Code references teacher mention of her relationship with student(s) during expansive probe section of cognitive interview.

### **Factor\_Informant\_Teaching Experience**

Teacher references personal teaching experience.

## Appendix D

### **Factor\_Informant\_Training**

Teacher references personal training experience.

### **Freq\_Attn\_O\_Attention Behaviors**

Rating based on frequency of teacher observing student physical paying attention behaviors (e.g. nodding, hand raised, materials ready)

### **Freq\_Attn\_O\_Effort**

Rating based on the frequency on teacher observation of student effort to pay attention

### **Freq\_Attn\_O\_Growth**

Rating based on student growth.

### **Freq\_Attn\_O\_Impresion of Awareness**

Rating based on teacher impression of student's level of awareness.

### **Freq\_Attn\_O\_NOT\_Attention Behaviors**

Rating frequency of student paying attention based on student growth from beginning of year.

### **Freq\_Attn\_O\_NOT\_Output**

### **Freq\_Attn\_O\_Output**

Rating based teacher observation of student verbal or written output following instruction or directions.

### **Freq\_Attn\_Q\_#Prompts**

Rating frequency of student paying attention based on # of prompts, reminders, redirections by teacher

### **Freq\_Attn\_Q\_#Subjects**

Rating frequency of student paying attention based on # of academic subjects that students typically attend to

### **Freq\_Attn\_Q\_%Time**

Rating frequency of student paying attention based on % of total time

### **Freq\_Attn\_Q\_Avg**

Rating based on average the amount of time student pays attention. This code is applied when no reference is made to the period that is included in the average.

### **Freq\_Attn\_Q\_Avg over Weeks**

Rating based on average amount student has paid attention over past few weeks

## Appendix D

### **Freq\_Attn\_Q\_Avg over Year**

Rating based on average amount student pays attention since beginning of year

### **Freq\_Attn\_Q\_Impression**

Frequency rating of attention based on teacher's general impression of student, or the way teacher imagines student behaving in classroom.

### **Freq\_Attn\_Q\_Time Sampled Varies by Student**

Frequency rating based on different time intervals for students with different levels of need.

### **Freq\_Group\_O\_Eagerness**

Frequency rating for contributing to group efforts based teacher assessment of student's eagerness and enthusiasm to participate in group effort.

### **Freq\_Group\_O\_Equitable Potential**

Frequency rating for contributing to group efforts based on the teacher assessment of whether student is contributing what he or she can contribute. Different expectations of different students.

### **Freq\_Group\_O\_Impact on Group**

Frequency rating for contributing to group efforts based on impact on group functioning. Students who are overly assertive, or who put up roadblocks would receive lower ratings than students who engage more collaboratively.

### **Freq\_Group\_O\_Inclusivity**

Frequency rating for contributing to group efforts based on the student's efforts to ensure that all members of group are included.

### **Freq\_Group\_O\_Leadership**

Frequency rating for contributing to group efforts based on teacher perception of student's confidence and leadership in group work. For teachers attending to leadership, more passive students would be judged as less frequently contributing to group work.

### **Freq\_Group\_O\_Physical Engagement**

Frequency rating for contributing to group efforts based on teacher observation of student body language (e.g. leaning in, eye contact with speaker) and physical acts (e.g. building structure).

### **Freq\_Group\_O\_Prompts**

Frequency rating for contributing to group efforts based on how frequently teacher needs to prompt student to engage.

## Appendix D

### **Freq\_Group\_O\_Quiet Challenge**

Teacher specifically references challenge with evaluating quiet students or the difficulties quiet students have contributing to group efforts.

### **Freq\_Group\_O\_Verbal Contribution**

Frequency rating for contributing to group efforts based on student's verbal contributions.

### **Freq\_Group\_Q\_Average**

Frequency rating for contributing to group efforts based on averaging experiences of student over time.

### **Freq\_Group\_Q\_Growth**

Frequency rating for contributing to group efforts based on teacher assessment of student growth over time.

### **Freq\_Group\_Q\_Recent Events**

Frequency rating for contributing to group efforts based on more recent events.

### **Freq\_Group\_Q\_Specific Events**

Frequency rating for contributing to group efforts based on recall of specific events.

### **Freq\_Nice\_O\_Growth**

Frequency rating based on teacher perception of student growth in student doing something nice for others.

### **Freq\_Nice\_O\_Negative Behaviors**

Teacher bases frequency of student doing something nice for others on the # of arguments or negative behaviors that she observes involving the student.

### **Freq\_Nice\_O\_Observable Behavior**

Frequency rating based on teacher's observation of specific actions that are visible or audible.

### **Freq\_Nice\_O\_Opportunities**

Frequency rating based on teacher's assessment of the # of opportunities that student recognizes and acts on to do something nice for somebody.

### **Freq\_Nice\_O\_Peer Relations**

Frequency rating based on teacher's observation of student's interactions with peers. Harmonious peer relations suggest that the student does nice things for others.

## Appendix D

### **Freq\_Nice\_O\_Prompting**

Teacher bases frequency of student doing something nice for others based on whether student performs acts independently or requires prompting.

### **Freq\_Nice\_O\_Scope**

Frequency dependent on who student is nice to. More favorable ratings for students who are nice to those outside of their friend group.

### **Freq\_Nice\_O\_Trait**

Frequency rating based on teacher's general experiences of child as a nice or kind child. Teacher explains student doing something nice as a trait.

### **Freq\_Nice\_Q\_# observed acts per day or week**

Frequency rating based on the the number of observable positive deeds. In some instances, teachers hold #deeds in particular time frame such as day or weeks. Negative positive acts coded separately.

### **Freq\_Nice\_Q\_Average**

Teacher bases frequency of student doing something nice on average of the experiences that she has had with student over some period.

### **Freq\_Nice\_Q\_Cummulative Experience**

Teacher bases frequency of student doing something nice for others on her cumulative experience of the student.

### **Freq\_Nice\_Q\_Recent Events**

Teacher bases frequency of student doing something nice for others on recent specific events.

### **Freq\_Nice\_Q\_Recognition may be happening but not observed**

Teacher specifies that students may be engaging in doing something nice for somebody, but that these deeds are not observed.

### **Freq\_Nice\_Q\_Specific Events**

Teacher bases frequency of student doing something nice for somebody on her recollection of specific events.

### **Gender\_Female**

Student identified as female.

### **Gender\_Male**

Student identified as male.

## Appendix D

### **Group\_Def\_Advancing Goal**

Definition includes advancing goal or specific outcome. Focus may be on product or 'success.'

### **Group\_Def\_Affect**

Definition includes demeanor, mood, or affect of student.

### **Group\_Def\_Equitable**

Definition mentions equitable contributions or describes the differences in student abilities to participate.

### **Group\_Def\_Including Others or Teamwork**

Definition includes student awareness of others through efforts to include others, work collaboratively, and engage in team work. Definition may address different roles.

### **Group\_Def\_Listening to Others' Ideas and Perspectives**

Definition includes listening to and considering the ideas and perspectives of others.

### **Group\_Def\_Not Speaking OK**

Definition specifies that students need not be verbally participating to be contributing to group efforts.

### **Group\_Def\_Participating Verbally or Physically**

Definition includes verbally or physical contributions to group activities.

### **Important Quote**

Quote flagged as important in understanding teacher thought processes

### **Nice\_Def\_Care without Act**

Definition includes awareness, care, and compassion for others without mention of a specific act.

### **Nice\_Def\_Ccomplimenting or Kind Words**

Definition includes verbal compliments or using kind words toward another.

### **Nice\_Def\_Elicit Pleasant Emotions**

Definition includes behaviors that elicits pleasant emotions in others. The behavior is not necessarily in response to a perceived need.

### **Nice\_Def\_Encouraging**

Definition includes the verbal encouragement of others.

## Appendix D

### **Nice\_Def\_Fills Bucket**

Definition refers to book entitled Have You Filled a Bucket Today by Carol McCloud.

### **Nice\_Def\_Nonverbal Behaviors**

Definition includes non-verbal behaviors that include, smiles, opening the door, eye-contact, putting arm around someone.

### **Nice\_Def\_Others First**

Definition mentions importance of selfless acts or putting others first.

### **Nice\_Def\_PandR\_Academic Need**

Definition refers to students perceiving and responding to academic need such as helping to redirect behavior toward learning or helping to learn content, or have materials in place.

### **Nice\_Def\_PandR\_SocEmo Need**

Definition mentions students perceiving and responding to a student's social or emotional needs such as doing something when student sad or socially not included.

### **Nice\_Def\_Self Initiated Act**

Definition references voluntary act that may be unexpected or unnecessary and shows an awareness and care for another person.

### **Nice\_Def\_Sharing**

Definition includes mention of sharing.

### **Nice\_Def\_Small Gestures**

Teacher references "small gestures" or "simple acts" of kindness.

### **Race\_African-American**

Teacher identifies student as Black or African-American.

### **Race\_Asian**

Teacher identifies student as Asian.

### **Race\_Immigrant**

Teacher identifies student as an immigrant.

### **Race\_Immigrant\_Eastern European**

Teacher identifies student as an immigrant from Eastern Europe.

## Appendix D

### **Race\_Immigrant\_Iceland**

Teacher identifies student as an immigrant from Iceland.

### **Race\_Immigrant\_Persian**

Teacher identifies student as an immigrant from Persia or Iran.

### **Race\_Immigrant\_Syrian**

Teacher identifies student as an immigrant from Syria.

### **Race\_Latino\_Hispanic**

Teacher identifies student as Latino or Hispanic

### **Race\_Multi-Ethnic**

Smart code that contains codes in which teacher identifies student as belonging to two racial ethnic groups.

### **Race\_Multi-Ethnic-Asian**

Teacher identifies student as belonging to two ethnic or racial groups with one being Asian.

### **Race\_Multi-Ethnic\_Black\_White**

Teacher identifies student as belonging to two ethnic or racial groups with one being Black or African-American.

### **Race\_Multi-Ethnic\_BlackCambodian**

Teacher identifies student as belonging to two ethnic or racial groups with one being Black and one being Cambodian.

### **Race\_Multi-Ethnic\_Japanese**

Teacher identifies student as belonging to two ethnic or racial groups with one being Japanese.

### **Race\_Multi-Ethnic\_Korean**

Teacher identifies student as belonging to two ethnic or racial groups with one being Korean.

### **Race\_Multi-Ethnic\_Philipino**

Teacher identifies student as belonging to two ethnic or racial groups with one being Philippine.

### **Race\_Multi-Ethnic\_Taiwanese\_White**

Teacher identifies student as belonging to two ethnic or racial groups with one being Taiwanese.

### **Race\_White**

Teacher identifies student as White or Caucasian.

## Appendix D

### **Response\_Averaging**

Teacher references differences in student behavior from day to day or across contexts and references these differences when selecting a response that takes the described variability into account.

### **Response\_Low**

Teacher reports tendency to rate her students on the lower end of frequency scale.

### **Response\_Mid Scale**

Teacher indicates likelihood of using moderate ratings.

### **Response\_Mid Scale\_EP**

Teacher indicates likelihood of using moderate ratings during expansive probe section of cognitive interview.

### **Response\_Plus or Minus**

Teacher rates student using plus or minus or mentions that actual rating would fall between two response options.

### **Response\_Revised\_Positive**

Teacher modified original rating to a more favorable rating

### **Response\_Uncertainty**

Teacher expresses uncertainty on which response option to select in rating frequency of student behavior.

### **Response\_Uncertainty\_EP**

Teacher expresses uncertainty in how to rate student behavior during expansive probe section of cognitive interview.

### **Services\_504**

Student has a 504 Plan.

### **Services\_ELL\_Current**

Student currently receiving ELL services.

### **Services\_ELL\_Exited**

Student was eligible for ELL services and has been exited.

### **Services\_Free Lunch**

Student receives free or reduced lunch services.

## Appendix D

### **Services\_Gifted**

Student receives services for gifted students.

### **Services\_SPED**

Student receives special education services.