ORIGINAL PAPER

Improving Classroom Quality with The RULER Approach to Social and Emotional Learning: Proximal and Distal Outcomes

Carolin Hagelskamp · Marc A. Brackett ·
Susan E. Rivers · Peter Salovey

Abstract The RULER Approach to Social and Emotional Learning ("RULER") is designed to improve the quality of classroom interactions through professional development and classroom curricula that infuse emotional literacy instruction into teaching–learning interactions. Its theory of change specifies that RULER first shifts the emotional qualities of classrooms, which are then followed, over time, by improvements in classroom organization and instructional support. A 2-year, cluster randomized controlled trial was conducted to test hypotheses derived from this theory. Sixty-two urban schools either integrated RULER into fifth- and sixth-grade English language arts (ELA) classrooms or served as comparison schools, using their standard ELA curriculum only. Results from multilevel modeling with baseline adjustments and structural equation modeling support RULER's theory of change. Compared to classrooms in comparison schools, classrooms in RULER schools exhibited greater emotional support, better classroom organization, and more instructional support at the end of the second year of program delivery. Improvements in classroom organization and instructional support at the end of Year 2 were partially explained by RULER's impacts on classroom emotional support at the end of Year 1. These findings highlight the important contribution of emotional literacy training and development in creating engaging, empowering, and productive learning environments.

C. Hagelskamp
Public Agenda, 6 East 39th Street, New York, NY 10016, USA

M. A. Brackett (✉) · S. E. Rivers · P. Salovey
Yale Center for Emotional Intelligence, Yale University,
P.O. Box 208376, New Haven, CT 06520-8376, USA
e-mail: marc.brackett@yale.edu

Introduction

Classroom interactions have a powerful influence on youth development. In high quality classrooms, interactions are respectful, youth are granted autonomy in the learning processes, teaching is organized, discipline is fair, and instructions are appropriate and effective (Catalano et al. 2004; Jennings and Greenberg 2009). Youth in these environments are most likely to perform well academically, develop strong social-emotional skills, and retain motivation to work and learn (Durlak et al. 2011). However, the majority of youth in the United States do not have this experience (Pianta et al. 2007). Many classroom environments fail to engage students academically, leave learning potential untapped, and provide little structured opportunity for building pro-social skills. In the worst cases, student–teacher conflict and peer bullying are high and academic failure common (Aud et al. 2010). Thus, increasing attention is being paid to intervention programs that target classroom interactions to foster social-emotional and academic well-being among youth (Durlak et al. 2011).

This paper reports on the effectiveness of one such intervention, The RULER Approach to Social and Emotional Learning ("RULER"), for improving classroom quality in fifth- and sixth-grade English language arts (ELA) classrooms over 2 years. RULER is a social and emotional learning (SEL) program that targets classroom interactions and instruction through teacher training and lessons that are infused into core curricula (Maurer and Brackett 2004). Utilizing data from a longitudinal, school-randomized

controlled trial (RCT), we first examine RULER's impact on a comprehensive assessment of classroom emotional, instructional, and organizational quality (i.e., scores on the Classroom Assessment Scoring System (CLASS); Pianta et al. 2008) after 2 years of program implementation. We then examine one process by which RULER may impact overall classroom quality, namely the hypothesis that RULER impacts a classroom's overall quality by first targeting the emotional climate, which in turn helps to enhance both classroom instructional and organizational quality.

Classroom Quality: A Multi-Dimensional Construct

To foster positive youth development, educational settings must meet students' multifaceted developmental needs (Catalano et al. 2004). Hence, classroom quality must be assessed across multiple dimensions, including social-emotional climate, instructional support, and classroom organizations (Hamre et al. 2007). Each dimension corresponds to a set of students' needs. The social-emotional climate of a classroom describes relationships between and among students and teachers as demonstrated by supportive and caring interactions, positive feelings, a shared sense of cohesion and respect, and teachers' sensitivity to student needs. Instructional support is expressed when teachers facilitate students' sense of accomplishment, scaffold learning, and give constructive feedback. Classroom organization refers to activities and disciplinary practices that keep classrooms predictable, efficient, and goal oriented (Hamre et al. 2007). When classrooms exhibit quality in all three domains, youth are more likely to have their social and emotional needs met and to achieve academically (National Research Council & Institute of Medicine 2002).

The relationship among the three domains of classroom quality is best understood as a system of transactional processes (Pianta and Allen 2008). In other words, the overall classroom climate is achieved and maintained through reciprocal and mutually reinforcing experiences in each domain. Nevertheless, theorists and practitioners vary with regard to which classroom quality domain they perceive to be the most impactful target of change. Some view classroom organization and instructional quality as the predominant drivers of classroom quality and student success (e.g., Rutter and Maughan 2002). Others, including the authors, argue that students' and teachers' social-emotional experiences and skills constitute the basic building blocks for a productive learning environment (Brackett and Katulak 2006; Jennings and Greenberg 2009). RULER's theory of change, described below, elucidates the argument for classroom social-emotional climate as the primary intervention target.

Changing Classroom Quality Through Social and Emotional Learning Programs

One evidence-based approach to influencing social-emotional climates in classrooms is to implement a SEL program. SEL programs are systematic educational strategies designed to cultivate skills including self- and social awareness, emotion regulation, responsible decision making, problem solving, and relationship management (Greenberg et al. 2003); yet individual SEL programs can vary in the focus of their skill-building efforts, e.g., some may focus on emotion regulation while others target coping or social support (CASEL 2003). A recent meta-analysis of 213 studies shows that students in schools that use universal SEL programs (programs that are implemented school-wide across the entire student body) have improved social, emotional, and academic skills (i.e., an 11-percentile point increase in achievement as indicated by grades and standardized test scores) and exhibit more pro-social behavior and less emotional distress (Durlak et al. 2011).

The RULER Approach to Social and Emotional Learning ("RULER")

RULER is a universal SEL program that targets five key emotion skills based on the achievement model of emotional intelligence (Rivers and Brackett 2011; Salovey and Mayer 1990). These skills include *recognizing* emotions in oneself and others, *understanding* the causes and consequences of emotions, *labeling* emotions with an accurate and diverse vocabulary, and *expressing* and *regulating* emotions in socially appropriate ways (Rivers and Brackett 2011). This achievement model of emotional intelligence proposes that emotional literacy—defined as having a mastery of the five RULER skills and appreciating the significance of these skills for social interactions, personal growth, and learning—is acquired through experience and the acquisition of emotion-related knowledge and skills. This acquisition ideally occurs in environments that are safe, supportive, and tolerant of all emotions; where consistent practice using the RULER skills is encouraged; and where feedback is provided by adults who also model the RULER skills. RULER targets emotions because they play a key role in social interactions, learning, self-reflective thinking, and perspective taking (Salovey and Mayer 1990). Accumulating evidence supports the link between the emotion skills described and individuals' social competence, psychological well-being, and academic performance (Fine et al. 2003; Rivers et al. 2012).

RULER is available for kindergarten through grade eight, with curricular units that extend across the academic year and instructional tools that are used daily. The approach combines comprehensive professional development for

school leaders and teachers with a literacy-based, skill-building SEL curriculum for students (the "Feeling Words Curriculum"). The Feeling Words Curriculum encourages educators and students to analyze the emotional aspects of personal experiences, academic materials, and current events. For example, in the unit on the feeling work "alienated", students might analyze a reading passage within their ELA curriculum to identify the causes and consequences of a character's feelings of alienation, then discuss with peers their own experiences with feeling alienated, and finally learn strategies for regulating feelings of alienation (Rivers and Brackett 2011). RULER's professional development component equips teachers with the knowledge, awareness, and skills necessary to build and maintain supportive relationships with students and encourage cooperation. It also teaches educators how to implement the lessons that comprise the Feeling Words Curriculum.

RULER's Theory of Change

Figure 1 illustrates RULER's theory of change. RULER has two inputs: (1) Teacher training and coaching in the RULER skills and the Feeling Words Curriculum and (2) the Feeling Words Curriculum itself. Those inputs have two proximal targets (listed to the left under Classroom-Level Change): (1) the quality of social-emotional interactions in the classroom, and (2) the emotional literacy skills of students and teachers. These outcomes mutually reinforce each other and increase the overall social-emotional climate of the classroom.

Next, RULER's theory of change specifies two distal outcomes on the classroom level: classroom instructional support and classroom organization (listed to the right under Classroom-Level Change). The model proposes that a positive social-emotional climate in the classroom is the basis for productive teaching and learning and effective classroom management. The theoretical rationale for this proposition is twofold. First, by definition, a positive social-emotional classroom climate meets youths' basic developmental needs for caring and supportive relationships coupled with the experience that their opinions count and are respected (Deci and Ryan 1985). When these needs are met, students are more motivated to learn (Rimm-Kaufman et al. 2007) and feel safe to engage with more challenging academic tasks and new material (Hamre and Pianta 2005). In turn, students are more receptive to teachers' instructions and expectations (Wentzel 2002), which further fosters a classroom environment that is conducive to high quality instruction and effective classroom organization.

Second, teachers, like their students, need to feel supported and competent in the classroom to be motivated, engaged, and perform well (Hakanen et al. 2006). The quality of relationships with students is among teachers' primary sources of job satisfaction and distress (Sutton and Wheatley 2003). Observer ratings of emotional support and student engagement in the classroom have been associated with teachers' beliefs in their agency and instructional competency (Malmberg and Hagger 2009). Moreover, teachers who are skilled at understanding and regulating their own emotions should be better at maintaining a regulated classroom and high quality instruction over time (Jennings and Greenberg 2009). A positive social-emotional climate in the classroom is thus expected to serve students and teachers and lead to improved classroom management and instructional practices.
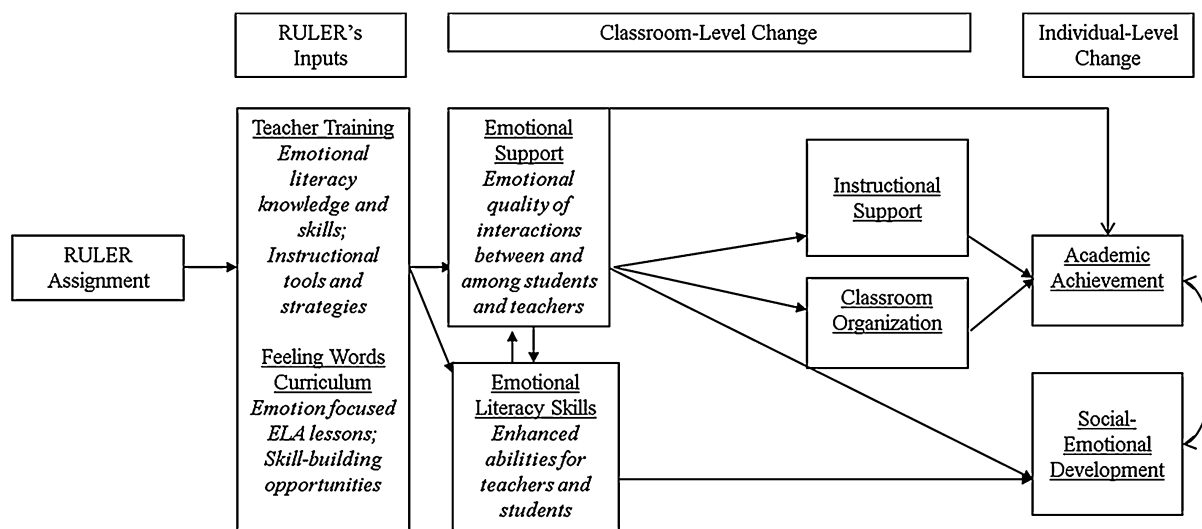

Fig. 1 Theory of change for The RULER Approach

Finally, RULER's theory of change specifies students' academic achievements and social-emotional behavior and well-being as distal, individual-level outcomes of the approach (listed under Individual-Level Change). These theoretical links are derived from the literature on the impact of classroom quality and SEL programs on youths' developmental outcomes, as reviewed earlier. They also have been supported by a previous, yet non-randomized study on the effectiveness of RULER. Students in fifth- and sixth-grade classrooms that integrated RULER performed better in writing and were better at following directions than students in comparison classrooms (Brackett et al. 2010).

The Present Study

The current paper focuses specifically on RULER's impacts on classroom emotional support, instructional support, and classroom organization. It utilizes data from the first cluster randomized control trial of The RULER Approach. As the first of its kind, this trial was limited to RULER's impacts in fifth- and sixth-grade classrooms over the course of 2 years. First year impact analyses have shown that after one school year of program implementation, independent observers rated fifth- and sixth-grade classrooms in intervention schools higher on emotional climate than classrooms in comparison schools (Rivers et al. 2013). The current paper seeks to extend these findings by examining RULER's longitudinal impact on all three domains of classroom quality (emotional support, instructional support, and classroom organization) after 2 years of program implementation, and by testing RULER's theory of change at the classroom-level, specifically its sequential impact on the three domains of classroom quality as indicated in Fig. 1. Two hypotheses are tested:

1. Classrooms in schools that implemented RULER exhibit significantly greater emotional support, higher instructional support, and better classroom organization after 2 years of program delivery compared to classroom in comparison schools.
2. RULER's earlier impacts on classroom emotional support, assessed at the end of the first year of program delivery, mediate RULER's later impacts on classroom instructional support and classroom organization by the end of Year 2.

Method

Participants

Sixty-two schools within the Roman Catholic Diocese of Brooklyn in Brooklyn and Queens, NY participated in the study. Schools were recruited into this study during a regularly scheduled, district-wide meeting of school principals in which the research team, in collaboration with the superintendent's office, presented the project. Approximately 70 of 109 principals were in attendance; 66 volunteered their fifth- and sixth-grade English language arts (ELA) classrooms to participate. Post randomization, four schools dropped out of the study (two from each condition), leaving 62 schools in the study. At baseline (spring 2008), these schools ranged in size from 186 to 557 students ($M = 325.24$, SD $= 98.47$) with the number of students per teacher ranging from 17 to 35 ($M = 24.35$, SD $= 3.69$). Ethnic/racial minority students made up between 5.80 and 100 % in each school ($M = 67.58$ %, SD $= 32.56$ %). The percentage of students receiving free lunch in each school ranged from 0 to 94 % ($M = 24.09$ %, SD $= 32.53$ %). The district office provided data from 2007/2008 fourth-grade students as a proxy for overall school academic performance. Schools ranged from 5.90 to 85.70 % of fourth-grade students performing poorly in ELA (Level 1 and 2; $M = 29.26$ %, SD $= 15.91$ %), and 0–66.60 % of fourth graders performing poorly in math (Level 1 and 2; $M = 20.87$ %, SD $= 14.91$ %). About half of the schools (53.20 %) reported already having a program that may target social-emotional skills; e.g., Learning for Life (www.learning-for-life.org).

Across these 62 schools, 164 classrooms and 90 teachers (91 % female) participated in the study at baseline. Teachers reported an average of 14.71 years of teaching experience (SD $= 10.08$, range 1–35) and having worked 9.27 years at their current school (SD $= 8.90$, range 1–34 years). Forty-three percent of teachers completed a master's degree, 15 % were working toward a master's degree, 33 % earned bachelor's degrees, and 8 % did not provide education information.

Randomization

To assign schools to conditions (RULER program or no-RULER comparison) all recruited schools were first randomly placed into pairs. Each pair was assigned either a 0 or 1 using a random numbers generator. Zero meant the first school of that pair was assigned to the comparison condition, and the other to the program condition; if a 1 was generated, the opposite assignment was made. To estimate whether randomization was effective, baseline data from program and comparison schools were compared for differences in the three outcome variables (emotional support, classroom organization, instructional support), teacher demographics (gender, years of teaching, years of at the school, education), and school characteristics (size, student–teacher ratio, percent free lunch, ethnic

composition, presence of another SEL program, fourth-grade school-averaged performance in math and ELA). Differences emerged on two variables. RULER schools were smaller in size (M [SD] $= 290.94$ [81.58]) than comparison schools (M [SD] $= 359.55$ [103.09]), $t(56.99) = 2.91$, $p = .005$. In RULER schools, classroom organization was lower (M [SD] $= 5.38$ [.13]) compared to comparison schools (M [SD] $= 5.68$ [.07]), $\gamma = -0.30$, SE $= 0.15$, $t = -1.98$, $p = .048$. Consequently, all analyses adjusted for baseline differences in school size and classroom organization.

Procedure

Timeline

Schools were recruited into the study in January 2008. Baseline data collection took place in April and May 2008; in July 2008 schools were assigned randomly to condition. In the first year of implementation (Year 1), RULER training commenced in October 2008 for program schools, and teachers were asked to begin using the intervention immediately and to continue until the end of the academic year. Post-baseline data were collected at four time points over two academic years: in Year 1 during fall 2008 (October/November) and spring 2009 (April/May); and in Year 2 during fall 2009 (October/November) and spring 2010 (April/May). Each wave of data collection occurred over an 8-week period. The current analyses focus on outcome data from the end of Year 2 (spring 2010) and include assessments taken at baseline and at the end of Year 1 as predictor variables.

Implementation of Teacher Training

Teachers in RULER schools were invited to attend approximately 9 h of RULER training across 2 days at the beginning of the first academic year of the study. They were then paired with a certified RULER coach with whom they met five times across each year. The training included didactic instruction, role-playing exercises, lesson design, observations of instructions, and feedback. Coaching sessions included sharing of successes and challenges, the provision of teaching tips and resources, and observations of lessons with feedback. Three optional booster training sessions were offered to teachers throughout the school year. Attendance records showed that 88.90 % of teachers attended the training session in Year 1, and 75.50 % of teachers attended at least four of the five coaching sessions offered to them ($M = 4.02$, SD $= 0.92$) in Year 1.

In Year 2, teachers in RULER schools again were offered training at the beginning of the year, followed by individual coaching sessions throughout the year.

Returning teachers were invited to attend the training and asked to attend at least two coaching session. Teachers new to RULER schools in Year 2 were invited to attend the training and asked to attend six coaching session throughout the year. In response to participants' feedback, booster sessions in Year 2 were replaced with three additional on-site training sessions, each focusing on a specific RULER activity. Attendance records showed that 97.91 % of fifth- and sixth-grade teachers attended the training session in Year 2. Among returning teachers, 90 % attended two or more coaching sessions; 10 % attended one coaching session. Among new teachers, 61.54 % attended at least five of the six coaching sessions offered ($M = 4.65$, SD $= 1.79$) in Year 2.

Implementation of the Curriculum

Fifth- and sixth-grade teachers at RULER schools were provided with 12 units of the Feeling Words Curriculum to be implemented across one academic year. Each unit focused on one feeling word, such as *excitement*, *shame*, *alienation*, and *commitment*, and included five lessons or "steps" each of which lasts approximately 15–20 min, yielding a total of 60 possible lessons. Teachers were asked to integrate the lessons into regular classroom instruction over the course of a 2-week period per unit. Across feeling word units, the five steps followed a basic structure. Briefly, teachers introduced the feeling word using a personalized connection (Step 1); students connected the feeling word to academic material or current events (Step 2); students demonstrated the meaning of the word through a visual or performing arts activity (Step 3); students discussed the feeling word with family members and wrote a short paragraph about the conversation (Step 4); and the class discussed and evaluated different strategies for managing the feelings that were part of family discussions, academic material, or current events (Step 5). Teachers were asked to implement one new unit every 2–3 weeks. Different sets of units were provided for each grade level. On average, teachers in RULER schools reported completing 7.38 (SD $= 2.79$) of the 12 units (each consisting of five lessons) in Year 1, and 7.17 (SD $= 3.72$) throughout Year 2.

Data Collection

The university's informed consent procedures were followed throughout the project. Each wave of data collection included observational assessments of the quality of interactions in the classroom. To collect this data, teachers used digital video equipment provided by the researchers and also received training and written instructions on how

to operate it. Teachers were asked to videotape their entire ELA class (about 40 min in length) on three different days of regular classroom activity (i.e., not on days in which tests were administered or movies were shown) within a 4- to 6-week period. Teachers returned videotapes using preaddressed, stamped envelopes.

Measures

Classroom Quality

The quality of interactions in the classroom was measured through observational assessments using the CLASS (Pianta et al. 2008). The CLASS was developed from extensive national, federally-funded, observational studies, as well as thorough reviews of relevant literature (Hamre et al. 2006). The CLASS has both face and construct validity, and predicts academic and social adjustment (Brackett et al. 2011; NICHD Early Child Care Research Network 2003). The CLASS is designed to distinguish between three domains of classroom quality: *emotional support*, *instructional support*, and *classroom organization*. Each domain consists of three to four dimensions. Emotional support is comprised of four dimensions: positive climate (e.g., warmth and respect in classroom interactions), negative climate (e.g., hostility in classroom interactions), teacher sensitivity (e.g., the degree to which teachers respond to students' emotional and academic needs), and teachers' regard for student perspective (e.g., the extent to which classroom activities incorporate students' points of view). Instructional support is comprised of three dimensions: concept development (e.g., the degree to which students are encouraged to engage in higher order thinking), quality of feedback (e.g., the use of feedback that extends student learning), and language modeling (e.g., the extent to which teacher–student conversations encourage students to use new words). Classroom organization also is assessed on three dimensions: behavior management (e.g., teachers' ability to monitor misbehavior), productivity (e.g., efficient transitions between activities), and instructional learning formats (e.g., the effectiveness by which materials are used to engage students). CLASS coders assigned a score to each dimension based on the presence or absence, frequency and quality of specific observable indicators using a 7-point scale (1–$2 = low$, 3–$5 = medium$, 6–$7 = high$).

At each wave of data collection, teachers were asked to submit three videotaped classroom sessions. Videotapes were converted into two segments of equal length, ranging from 8 to 20 min. Few segments (i.e., 2.85 % in Year 2) needed to be discarded as they contained insufficient data to code (e.g., were shorter than 8 min, had poor audio quality, or students were not visible). Coders employed the CLASS to rate each segment of classroom interactions on each of the 10 dimensions. To maintain reliability and prevent coding drift, several steps were taken. Coders completed a 2-day CLASS certification training and were required to pass initial and weekly reliability testing. If a coder did not demonstrate reliability (80 % of assigned codes within one point of the master code), CLASS master trainers worked with the coder until reliability was achieved and maintained. A total of 33 unique coders contributed to the coding across the project period. Segments were assigned randomly to one or more coders who were blind to the classroom's condition assignment. For the spring of Year 2, an average of 4.44 segments (SD = 1.90) were coded for each videotaped classroom, ranging from one to six segments per classroom (excluding classroom for which no videotape was submitted, see missing data section below). Each segment was coded between one and three times by a unique coder. The vast majority of segments were coded by two unique coders (84.53 %) As opposed to other studies in which the majority of data is coded by only one trained coder or in-classroom observer (Hamre et al. 2007), this data set allowed for inter-coder reliability checks across nearly all coded segments. A one-way random effects ANOVA was used to calculate the reliabilities of segment means furnished from the ratings of multiple independent coders (ICC($1,K$), (Shrout and Fleiss 1979). The resulting ICCs ranged from .47 to .70 across the 10 dimensions of the CLASS, indicating moderate reliability between any two coders. Aggregation of scores across coders was thus justified (LeBreton and Senter 2008).

To obtain one score for each CLASS dimension per classroom, we first averaged scores across all raters for a segment, and then aggregated scores across all segments. Intersegment reliabilities (Cronbach's α) ranged from .58 to .81 for the 10 CLASS dimensions. Finally, domain scores were created by averaging across the dimensions that represent emotional support ($\alpha = .82$), instructional support ($\alpha = .89$), and classroom organization ($\alpha = .75$).

The same coding and scoring procedures were employed to obtain CLASS scores from videotapes submitted at baseline and Year 1. Reliability coefficients for those time points were comparable to the ones reported for Year 2 CLASS scores and are available from the authors. For the purpose of this paper, classroom-level scores at baseline and at the end of Year 1 were aggregated within each school to obtain school-level indicators of baseline instructional support (ICC = .22) and classroom organizations (ICC = .52), and for emotional support at the end of Year 1 (ICC = .32). As discussed below, these school-level averages needed to be obtained in order to examine relationships across three academic years.

Analytic Plan

To test Hypothesis 1, we estimate the impact of RULER on each of the three domains of classroom quality (emotional support, instructional support, and classroom organization), using an intent-to-treat analysis and a multilevel modeling framework that accounts for the nesting of classrooms within schools. The intent-to-treat analysis obtained an unbiased estimate of the treatment effect by computing the average difference in classroom quality outcomes resulting from offering RULER to schools versus not (Bloom 2005). Multilevel or hierarchical linear modeling (HLM) allows for the simultaneous estimation of variation associated with classroom (level 1) and school-level (level 2) components while taking into consideration homogeneity of variance of residuals, yielding less biased parameter estimates (Raudenbush and Bryk 2002). RULER's impact on each classroom quality domain was estimated separately. All models were analyzed using MPlus6 (Muthen and Muthen 2010). Each model included grand-mean centered baseline covariates on which RULER and comparison schools differed (school size, classroom organization) and those that were related to biases in missing data (proportion of students per school who received free lunch), as described in the missing data analysis section below. All models used full information maximum likelihood (FIML) procedures. FIML provides efficient statistical parameter estimation from incomplete data and thus allowed for the retention of cases with missing data (Graham 2009). Effect sizes were calculated using Hedge's g (Hedges 2007). To estimate the amount of variance in each classroom quality domain that was uniquely accounted for by RULER, we compared each covariate model before and after the treatment variable (RULER versus comparison) was included and obtained a ΔR^2 statistic.

Further, we tested Hypothesis 2, namely RULER's theory of classroom-level change and the notion that RULER's impacts on instructional support and classroom organization at the end of Year 2 (i.e., Hypothesis 1) are mediated by its earlier impacts on classroom emotional support, as assessed at the end of Year 1. To this end, we employed structural equation modeling (SEM) and estimated indirect effects in a path model that specified direct paths from (1) RULER to Year 2 instructional support, (2) RULER to Year 1 emotional support, and (3) Year 1 emotional support to Year 2 instructional support. Figure 2 illustrates the path model. SEM estimates all paths in this model simultaneously and provides empirical estimate of the size and significance of the indirect effect of RULER on Year 2 instructional support via Year 1 emotional support. Bootstrapping methods were used to estimate the standard errors for the indirect effect. Bootstrapped standard errors result in non-symmetric confidence intervals

and take into account that the sampling distributions of indirect effects are likely to be skewed (Shrout and Bolger 2002). When the 95 % confidence bounds do not include zero, we can conclude with 95 % confidence that the indirect effect is not zero. Further, we calculated the *mediation ratio* (P_M), i.e., the ratio of the indirect effect over the total effect, to estimate the relative extent to which earlier impacts on emotional support mediate RULER's later impacts on instructional support (Huang et al. 2004). The same analytic procedures were employed to evaluate RULER's theory of classroom-level change with Year 2 classroom organization as dependent variable. Baseline school size, classroom organization, and proportion of students receiving free lunch served as covariates. Analyses were conducted in Mplus6 (Muthen and Muthen 2010). FIML procedure was employed to retain cases with missing data.

These SEM analyses relied on school-aggregated indicators of classroom level quality because relationships were tested across academic years. By utilizing fifth- and sixth-grade data from three consecutive spring semesters, no class could contribute to more than two time-points. Teachers from across the 62 schools in the study did not consistently follow classes (i.e., the same group of students) into the next year, and a classroom's student composition, too, changed between academic years. Therefore, the hypothesized process of change could not be examined within the same classrooms. Instead the analysis is based on the assumption that as schools (the level targeted by the intervention) are exposed to RULER over time, RULER effects will be identifiable in school-average levels of classroom quality indicators. By utilizing data from the same season (i.e., spring) over 3 years, seasonal biases were controlled for, such as the possibility that classroom quality drops in the last few months of every school year (Hamre et al. 2009).

We recognize that we cannot draw decisive causal conclusions from these indirect effects (or mediation) models. As many applied experiments, this study was
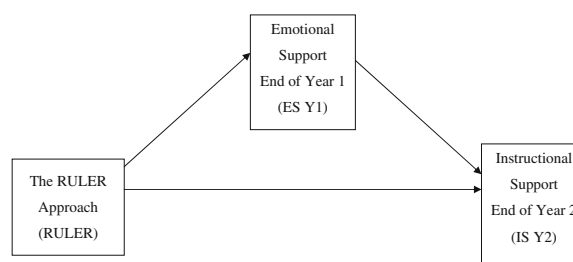

Fig. 2 Path model illustrating the hypothesized indirect effect of The RULER Approach on classroom instructional support after 2 years of program implementation via its earlier impacts on classroom emotional support after 1 year of program implementation

designed to randomize units (i.e., schools) to either program or comparison conditions, but could not also randomly assign units to the proposed mediator (i.e., Year 1 classroom emotional support). Mediation analyses with un-manipulated mediators cannot rule out the existence of unobserved variables that affect both the mediator and the outcome, which renders causal inferences impossible (for a detailed account this argument see Bullock et al. 2010). As such, this RCT allows us to draw causal inferences with respect to the interventions' direct effects on each domain of classroom quality (Hypothesis 1). However, causal relationships between earlier emotional support and later instructional support and classroom organization, and hence mediation, may only be suggested. Nevertheless, we took recommended measures to reduce inherent biases from our mediation models by including baseline measures of instructional support and classroom organization, respectively, as covariates (Cook et al. 2009). Including a pre-test variable has been shown to reduce biases in treatment effects by an average of 64 % in studies with education-related outcomes (Steiner et al. 2010).

Results

Preliminary Analyses

Sample Attrition

From the 62 schools that continued to participate after randomization, 57 schools remained in the study at the end of Year 2; the other five schools (three from the treatment condition and two from the comparison condition) had closed. At baseline, these five schools had a slightly lower student–teacher ratio (t (17.99) = 3.83, p = .001) and greater representation of ethnic/racial minority students (t (16.56) = −2.71, p = .032) compared to schools that stayed in the study. These school characteristics, however, were unrelated to classroom quality indicators at baseline. Moreover, these five schools did not differ from other schools with respect to school size, and indicators of SES and school performance (percent of student receiving free lunch, fourth-grade math and ELA performance). Data loss due to school attrition by Year 2 was thus treated as missing-at-random with respect to classroom quality. These schools were retained in the analyses using full information maximum likelihood procedures.

At the end of Year 2, 144 fifth- and sixth-grade classrooms taught by 96 teachers (85 % female) participated in the study. Most of these teachers (57 %) had participated in the study at baseline; 13 % entered the study in Year 1; 30 % were new to the study in Year 2. This turnover pattern reflects the typical turnover rate of teachers in urban

schools (Marinell and Pallas 2011), and it did not differ across RULER and comparison schools (χ^2 (2) = 0.53, p = .767). Teachers who participated only at baseline and those who entered the study in Year 2 did not differ with regards to education level, the number of years they had taught, or gender.

Missing Data on the Outcome Variables

At the end of Year 2, videotaped classroom observations were received from 93 of the 144 participating classrooms (65 % return rate). Sixty-seven teachers returned at least one tape (70 % return rate). Overall, observational data were obtained from 45 schools (79 % return rate for Year 2 participating schools). Return rates were comparable across RULER and comparison schools (χ^2 (1) = 1.87, p = .171).

Binary logistic regression analysis was used to predict missingness from school-level variables, including school size, student–teacher ratio, racial/ethnic composition, proportion of students receiving free lunch, presence of another SEL program, fourth-graders' math and ELA performance, as well as baseline school-level averages in observer-rated classroom emotional support, instructional support and classroom organization. One variable was associated with missing observational data; schools with a lower proportion of students receiving free lunch at baseline were more likely to miss observational data (b = −0.06, SE = .03, p = .041, OR = 1.07, 95 % CI [1.00, 1.13]). As this school characteristic also was associated with classroom emotional support at baseline (r (45) = −.45, p = .002), all analyses adjusted for the proportion of students in each school who received free lunch.

Because teachers determined whether or not their classrooms were videotaped, teacher characteristics (gender, education, years of teaching, years of at the school, and number of fifth- and sixth-grade classrooms taught in Year 2) also were examined as predictors of not returning videotapes. Using two-level linear regression models, adjusting for the fact that teachers were clustered in schools, no differences on these variables were found between teachers who had returned at least one videotape during the end of Year 2 data collection effort and those who had not. Finally, missing observational data was unrelated to students' ratings of the quality of their relationship with their ELA teachers at the end of Year 2, as assessed with Cook et al. (1995) *Affiliation with Teacher* measure in student surveys collected for the larger study. As this measure can be considered a contemporaneous indicator of the quality of interactions in classrooms, the fact that it did not predict missingness in observational data suggests that whether or not a classroom was videotaped at the end of Year 2 was unrelated to the quality of interactions in the classroom.

Missing Data on Predictor Variables

At baseline, 45 of the 62 schools had observational data from at least one classroom (72.58 % return rate). Binary logistic regression analysis showed that schools with a lower percentage of ethnic/racial minority students were more likely to miss baseline observational data ($b = 0.02$, SE = .01, $p = .049$, OR = 1.01, 95 % CI [1.00, 1.04]), but baseline percent of ethnic/racial minority students was unrelated to baseline observer-rated instructional support (r (45) = .01, $p = .963$) and classroom organization (r (45) = −.10, $p = .524$). At the end of Year 1, video-tapes were received from 32 of the 62 schools (51.6 % return rate), but the two groups of schools did not differ on school-level characteristics.

In sum, missing data patterns in this sample, although relatively substantial for some variables, were largely missing at random, justifying the use of full information maximum likelihood procedures (Graham 2009). With the exception of the proportion of students receiving free lunch, no other variables were related to attrition or missing data, or were associated with the predictor or outcome variables of interest in this paper. Therefore, no other covariates were included in the main analyses (Raudenbush et al. 2007). As a robustness check, we reran this paper's main analyses including all variables related to missingness (i.e., including minority representation and student–teacher ratio). The results were comparable to those presented below and are available upon request.

Correlations Between Classroom Quality Domains

Table 1 shows zero-order correlations among the outcome variables in this study. The three domains of the CLASS were highly correlated, but not completely overlapping. Because RULER's theory of classroom-level change has distinct hypotheses about shifts in these domains, the domains were not combined into a single factor. Moreover,

Table 1 Descriptive statistics, interclass correlations, and zero-order correlations between three domains of classroom quality after 2 years of program implementation (N = 144 classrooms in 62 schools)

	α	ICC	M	SD	1	2	3
1. Emotional support	.93	.44	4.62	0.50	1		
2. Instructional support	.74	.43	2.30	0.54	.71**	1	
3. Classroom organization	.80	.55	5.56	0.47	.67**	.60**	1

α = Reliability between coded segments from the same classroom (based on four segments per classroom). These ICCs estimate the proportion of variance due to differences between schools versus differences between classrooms

** $p < 0.01$

Hamre et al. (2007), drawing on data from over 4,000 US classrooms, provide empirical evidence that the three domains are indeed typically distinct.

Impact of RULER on Classroom Quality at Year 2

The first hypothesis tested was that classrooms in schools that implemented RULER exhibit greater emotional support, higher instructional support, and better classroom organization after 2 years of program delivery compared to classroom in comparison schools. Table 2 summarizes the results of the HLM analyses. Consistent with the theoretical model and our hypothesis, we found that by the end of Year 2, classrooms in RULER schools were rated significantly higher on emotional support, instructional support and classroom organization. Differences between RULER and comparison schools ranged from .48 of a standard deviation in the emotional support domain to .71 of a standard deviation in the instructional support domain. Those differences are interpreted conventionally as moderate to large effects (Cohen 1988). Correspondingly, RULER explained substantial amounts of variance in each classroom quality indicator, ranging from 15 % in the emotional support domain to 31 % in the instructional support domain—as indicated by the ΔR^2 statistic in the far right column of Table 2.

Evidence for RULER's Theory of Classroom-Level Change

The second hypothesis tested was that RULER's impacts on classroom instructional support and classroom organization by the end of Year 2 (as evidenced through our test of Hypothesis 1) are mediated by its earlier impacts on classroom emotional support, which was assessed after 1 year of program delivery. Table 3 summarizes the results of the SEM path analyses and their respective indirect effects estimates. Turning first to the results of the instructional support model, this model, adjusted for baseline covariates, fit the data well (χ^2 (4) = 3.21, $p = .524$, CFI = 1.00, RMSEA = .000). Two of the three paths of interest were significantly different from zero. First, RULER increased classroom emotional support after 1 year of program delivery (RULER $\Rightarrow$ ES Y1)—for more detail on RULER's Year 1 effects see Rivers et al. (2013). Second, and in alignment with Hypothesis 2, higher levels of Year 1 emotional support predicted more instructional support at the end of Year 2 (ES Y1 $\Rightarrow$ IS Y2). The direct path from RULER to Year 2 instructional support (RULER $\Rightarrow$ IS Y2) was not statistically significant with end-of-Year 1 emotional support in the model. Finally, and in alignment with RULER's theory of classroom-level change, RULER's indirect effect on Year 2 instructional

Table 2 Estimates of the impact of The RULER Approach on three domains of classroom quality after 2 years of program implementation (N = 144 classrooms in 62 schools)

	Adjusted means (SE)		b (SE)	95 % Confidence interval	p	Effect size (Hedge's g)	ΔR^2
	Ruler	Comparison					
Emotional support	4.78 (0.09)	4.50 (0.08)	0.24 (0.12)	[0.01, 0.47]	.043	.48	.15
Instructional support	2.47 (0.10)	2.11 (0.08)	0.36 (0.13)	[0.11, 0.61]	.005	.71	.31
Classroom organization	5.66 (0.07)	5.40 (0.10)	0.26 (0.12)	[0.03, 0.49]	.026	.56	.24

Impacts were estimated in a two-level regression model, adjusting for the nested structure of the data (classrooms in schools), and estimating a random effect at the school level. Models adjusted for baseline differences between RULER and comparison schools in school size, classroom organization (aggregated to the school level), and for the proportion of students receiving free lunch in each school. All covariates were grand-mean centered. Full information maximum likelihood estimation methods were employed to account for missing data. ΔR^2 is the difference in the amount of variance explained in each classroom quality domain between the covariate-only model and the model that added the treatment variable (RULER versus control)

Table 3 Results for SEM analyses estimating the indirect effect of The RULER Approach on change in classroom instructional support and classroom organization at the end of Year 2 via earlier impacts on classroom emotional support at the end of Year 1 (N = 62 schools)

Model	b	SE	p	95 % Confidence interval
Instructional support[a]				
RULER $\Rightarrow$ ES Y1	0.45	0.17	.006	[0.12, 0.79]
ES Y1 $\Rightarrow$ IS Y2	0.80	0.22	.000	[0.36, 1.25]
RULER $\Rightarrow$ IS Y2	0.07	0.12	.562	[−0.18, 0.32]
Indirect effect	0.36			[0.08, 0.80]
Total effect	0.43			[0.13; 0.75]
Classroom organization[b]				
RULER $\Rightarrow$ ES Y1	0.36	0.14	.014	[0.07, 0.64]
ES Y1 $\Rightarrow$ CO Y2	0.59	0.12	.000	[0.35, 0.82]
RULER $\Rightarrow$ CO Y2	0.05	0.08	.572	[−0.12; 0.21]
Indirect effect	0.21			[0.03, 0.39]
Total effect	0.25			[0.07, 0.46]

Models adjusted for baseline differences in school size, proportion of students receiving free lunch and classroom organization between RULER and comparison schools

[a] The instructional support model also adjusted for baseline instructional support. Model fit: χ^2 (4) = 3.21, p = .524, CFI = 1.00, RMSEA = .000

[b] Model fit: χ^2 (3) = .78, p = .854, CFI = 1.00, RMSEA = .000. The 95 % confidence intervals for the indirect and direct effects were obtained by the bias-corrected bootstrap with 1,000 resamples. RULER = The RULER Approach; ES Y1 = classroom emotional support after 1 year of program delivery; IS Y2 = classroom instructional support after 2 years of program delivery; CO Y2 = classroom organization after 2 years of program delivery; $\Rightarrow$ = affects

support via earlier shifts in emotional support was positive and significant—its 95 % bootstrapped confidence interval excluded zero. The indirect effect made up approximate 84 % of RULER's total effect on classroom instructional support ($P_M = 0.836$).

Now turning to the results of the classroom organization model, this model fit the data well (χ^2 (3) = .78, p = .854,

CFI = 1.00, RMSEA = .000). Consistent with Hypothesis 2, greater emotional support at the end of Year 1 predicted significant improvements in classroom organization at the end of Year 2 (ES Y1 $\Rightarrow$ CO Y2). Moreover, the estimated indirect effect of RULER on Year 2 classroom organization via earlier shifts in emotional support was positive and its 95 % bootstrapped confidence interval excluded zero. This effect accounted for approximately 81 % of RULER's total effect on classroom organization ($P_M = 0.814$). RULER's direct effect on end-of-Year 2 classroom organization was not statistically significant with end-of-Year 1 emotional support included in the model (RULER $\Rightarrow$ CO Y2).

Discussion

Building on theory and previous research, this study provides evidence that The RULER Approach affects the emotional, instructional, and organizational quality of fifth- and sixth-grade classrooms after 2 years of program implementation. Independent observers, blind to the experimental condition, rated classrooms in schools randomly assigned to RULER, on average, more favorably on all three domains of classroom quality compared to classrooms that were assigned to the comparison condition and received treatment as usual. These findings extend previous findings by showing that RULER's Year 1 impact on social-emotional processes in the classroom (i.e., emotional support) was sustained over 2 years of program implementation, and that after prolonged implementation, RULER's impact on classroom quality broadened to include positive effects on classrooms' instructional quality and organization.

This study tested not just the impact of an SEL program on multiple components of classroom quality, but also the process by which the program affects overall classroom quality. Aligning with RULER's theory of change, schools' average classroom emotional support assessed after the first

year of delivery predicted improvements in instructional support and classroom organizational after the second year of program delivery. Moreover, we provided evidence in support of our hypothesis that RULER's impacts on instructional support and classroom organization at the end of Year 2 were facilitate by it earlier impacts on classroom emotional support at the end of Year 1.

These results are particularly noteworthy given the methodological strengths of this study. Using data from a school RCT, the analyses provided a rigorous test of the impact of RULER on the classroom setting. RCTs control for demographic variation and related socio-economic factors that may contribute to outcomes. Analyses also adjusted for baseline differences in school-level demographics and outcome variables. Such experimental tests are critical for establishing the evidence base that SEL programs like RULER can enhance learning environments.

Further, this study employed the CLASS, the most widely used and validated tool currently available for the measurement of classroom quality in elementary and middle schools (Hamre et al. 2007), and relied on extensively trained independent coders who were blind to condition assignment; hence reducing potential biases of self-report assessments. Another strength is that the average number of RULER units and lessons taught (i.e., program dosage) remained stable and considerably high over 2 years of program delivery, with an average of about thirty-five lessons taught per classroom each year. This is a sign of consistent implementation fidelity among teachers over time and likely to be due to a number of factors, including RULER's reliance on well-trained, professional coaches who typically build strong working relationships with teachers. Furthermore, the program and the research were implemented with strong endorsement from the school superintendent and the principals who volunteered their schools. RULER may also naturally create dynamics within schools by which teachers increasingly support and motivate each other to remain faithful to the program over time. More research, however, is needed to understand teacher's fidelity to the program not only in terms of the number of lessons delivered, but also with regards to how well teachers implement each lesson and provide emotion skill building opportunities for students. Moreover, subsequent research needs to investigate how variation in implementation fidelity between teachers (or schools) moderate program impacts on classroom level outcomes (e.g., Reyes et al. 2012).

Finally, with the use of longitudinal data, structural equation modeling, baseline controls of the outcome variables, and bootstrapping methods for the estimation of indirect effects, this study provides a sophisticated test of the potentially causal process by which RULER first impacts classroom emotional quality, followed by impacts on instructional quality and classroom organization. Even though the analysis does not allow for the inference that these sequential changes (i.e., indirect effects) are necessarily causal (Bullock et al. 2010), it provides strong initial evidence for an as yet untested theory of classroom quality change.

Emotional Literacy and Classroom Quality

This study makes two important contributions to the conceptual understanding of the relationships between SEL and classroom quality. First, it supports the notion that SEL programs can be powerful initiatives to improve the quality of learning environments. In fact, the pattern and size of RULER's effects on emotional and instructional supports found in this study were remarkably similar to those found with an RCT of the 4Rs Program (Brown et al. 2010). In both studies, SEL programming yielded a moderate impact on emotional support and a large impact on instructional quality, as assessed through the CLASS. More specifically, the findings of this study support the central premise of RULER, namely that a focus on emotions and emotional literacy skills constitutes one promising means towards changing classroom quality. RULER's program components encourage frequent social interactions with peers and teachers, along with activities that focus on creativity, group problem solving, conflict resolution and empathy (Rivers and Brackett 2011). These program inputs are expected to contribute to an overall more positive classroom climate, as observed in more caring and sensitive interactions, more productivity and discipline, and higher quality of feedback and instruction. In future studies it will be important to directly compare the effects of RULER to those of other SEL programs, in order to identify the unique effect of RULER's focus on emotional literacy vis-à-vis that of alternative theory-driven program inputs.

Second, the results of this study advance understanding of the dynamics within the multidimensional system of classroom quality, and where to effectively intervene in the system. Having linked RULER's earlier impacts on classroom emotional support to later improvements in instructional support and classroom organization, this study is the first to suggest that the social-emotional qualities of a classroom are particularly effective targets to produce shifts in the classroom quality system as a whole. These findings align with SEL theories, including RULER's, which build on the notion that the emotional qualities of classroom interactions are the basis for productive teaching and learning (e.g., Hamre and Pianta 2005; Jennings and Greenberg 2009). The model proposed and tested herein further aligns with evaluation studies that most consistently report improvements in the emotional qualities of classrooms as a result of SEL programming (Raver et al. 2008; Solomon et al. 1996). This is not to say that social-emotional qualities are *necessarily* the precursor to higher

quality instruction and organization. It is possible that programs which primarily target a different domain of classroom quality (e.g., discipline, instructional quality), or those that combine program inputs that directly target all three domains of classroom quality, are equally effective in improving classroom quality (Rimm-Kaufman et al. 2007). At the very least, however, our findings support the conclusion that a program that strives to provide opportunities for cooperative and empathetic interactions in the classroom, and to develop emotional literacy skill in children as well as their educators, is an effective way to improve classroom quality overall, and that this overall change seems to be initiated through improvements in the social-emotional qualities of interactions between and among teachers and students in the classroom.

Limitations and Future Directions

Despite considerable methodological strengths, these results need to be interpreted in a context of some limitations. First, we tested RULER's theory of classroom quality change on the school level with a 1-year time lag between the proposed cause (shifts in emotional climate) and its effects (shifts in instructional support and classroom organization). As such this paper is informative about the process of change as whole schools adopt an SEL intervention over 2 years. This analysis, however, is less informative about the process of classroom-level change as it unfolds within individual classrooms. Even though each classroom was assessed twice within an academic year (i.e., fall and spring), the two assessments are insufficient to estimate the proposed causal process by which a shift in emotional support is expected to both appear over the course of the first weeks and months of RULER delivery and precede shifts in instructional quality and classroom organization later in the year. In future studies, monitoring of classrooms over several months within one academic year will allow us to test these causal processes.

Second, and related to the previous point, this study tested a sub-set of the components that comprise RULER's theory of classroom-level change. We thus view this study as an important first step to understanding the dynamics by which RULER impacts classroom overall quality. Further analysis must examine RULER's impacts on students' (and teachers') emotional literacy skills and assess the extent to which increases in these skills among students (and a teacher) within one classroom may partly contribute to improvements in overall classroom quality over time. Again, such analyses are conducted ideally with data that assess RULER's impacts more frequently and can thus track the process by which the program leads to sequential improvements within classrooms over the course of one school year.

Third, as documented, missing data patterns in this study were predicted by a number of school-level characteristics, such as the proportion of students receiving free lunch, student–teacher ratio, and racial/ethnic composition. Even though there was no evidence that these school characteristics were related to the quality of classrooms (and the analyses included adjustments for the ones that were), it is possible that the data do not generalize to schools at the extremes of these characteristics. Moreover, it is possible that non-measured factors that caused attrition biased these results. In addition, future research needs to determine whether the results of this study generalize to different types of schools than the ones in this study, such as to schools that are located in rural or suburban areas, to non-catholic schools, and to public schools.

Conclusion

To accomplish an educational agenda that addresses the needs of the whole child, school-based programming should strive to improve the quality of the environments in which academic, social, and emotional learning occurs (Greenberg et al. 2003). SEL programs like RULER provide this opportunity. Evidence-based SEL curricula both directly and indirectly address social and emotional competences across the curriculum. In addition to boosting academics, SEL programs intend to teach students the broad range of skills needed to cultivate quality relationships, and be psychologically and physically healthy. Yet, their success is dependent upon the extent to which learning occurs in caring, supportive, and empowering settings, and amplified by researchers' increasing capacity to conceptualize and assess setting-level dynamics (Seidman 2012). The present study provides evidence that a classroom's social-emotional quality constitutes an effective primary target of change when comprehensive improvement in classroom quality is sought. Moreover, it suggests that a specific focus on emotional literacy training is a promising avenue for such change, and it implies that emotional literacy skills may be a central component of the overall social-emotional quality of classrooms.

References

Aud, S., Hussar, W., Planty, M., Snyder, T., Bianco, K., Fox, M. A., et al. (2010). The condition of education 2010. NCES 2010-028. *National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education*. Washington, DC: U.S. Government Printing Office.

Bloom, H. S. (2005). *Learning more from social experiments: Evolving analytic approaches*. New York: Russell Sage Foundation.

Brackett, M. A., & Katulak, N. A. (2006). Emotional intelligence in the classroom: Skill-based training for teachers and students. In J. Ciarrochi & J. D. Mayer (Eds.), *Applying emotional intelligence: A practitioner's guide* (pp. 1–27). New York: Psychology Press/Taylor & Francis.

Brackett, M. A., Reyes, M. R., Rivers, S. E., Elbertson, N. A., & Salovey, P. (2011). Classroom emotional climate, teacher affiliation, and student conduct. *Journal of Classroom Interaction, 46*(1), 27–36.

Brackett, M. A., Rivers, S. E., Reyes, M. R., & Salovey, P. (2010). Enhancing academic performance and social and emotional competence with the RULER Feeling Words Curriculum. *Learning and Individual Differences*. doi:10.1016/j.lindif.2010.10.002.

Brown, J. L., Jones, S. M., LaRusso, M. D., & Aber, J. L. (2010). Improving classroom quality: Teacher influences and experimental impacts of the 4rs program. *Journal of Educational Psychology, 102*(1), 153.

Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what's the mechanism? (Don't expect an easy answer). *Journal of Personality and Social Psychology, 98*(4), 550.

CASEL [Collaborative for Academic, Social, and Emotional Learning]. (2003). *Safe and sound: An educational leader's guide to evidence-based social and emotional learning (SEL) programs.* Chicago, IL: Author.

Catalano, R. F., Berglund, L., Ryan, J. A. M., Lonczek, 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. doi:10.1177/0002716203260102.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale: Lawrence Erlbaum.

Cook, E. T., Greenberg, M. T., & Kusche, C. A. (1995). *People in my life survey: Attachment relationships in middle childhood.* Paper presented at the Society for Research in Child Development, Indianapolis, IN.

Cook, T. D., Steiner, P. M., & Pohl, S. (2009). How bias reduction is affected by covariate choice, unreliability, and mode of data analysis: Results from two types of within-study comparisons. *Multivariate Behavioral Research, 44*(6), 828–847.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior.* New York: Plenum.

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. doi:10.1111/j.1467-8624.2010.01564.x.

Fine, S. E., Izard, C. E., Mostow, A. J., Trentacosta, C. J., & Ackerman, B. P. (2003). First grade emotion knowledge as a predictor of fifth grade self-reported internalizing behaviors in children from economically disadvantaged families. *Development and Psychopathology, 15*(2), 331–342. doi:10.1017/S095457940300018X.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*, 549–576.

Greenberg, M. T., Weissberg, R. P., O'Brien, M. U., Zins, J. E., Fredericks, L., Resnik, H., et al. (2003). Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. *American Psychologist, 58*, 466–474. doi:10.1037/0003-066X.58.6-7.466.

Hakanen, J. J., Bakker, A. B., & Schaufeli, W. B. (2006). Burnout and work engagement among teachers. *Journal of School Psychology, 43*(6), 495–513.

Hamre, B. K., Mashburn, A. J., Pianta, R. C., Locasle-Crouch, J., & La Paro, K. M. (2006). *CLASS: classroom assessment scoring system, technical appendix.* Charlotteville, VA: University of Virginia.

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

Hamre, B. K., Pianta, R. C., & Chomat-Mooney, L. (2009). Conducting classroom observations in school-based research. In L. M. Dinella (Ed.), *Conducting science-based psychology research in schools* (pp. 79–105). Washington, DC: American Psychological Association.

Hamre, B. K., Pianta, R. C., Mashburn, A. J., & Downer, J. T. (2007). Building a science of classrooms: Application of the CLASS framework in over 4,000 US early childhood and elementary classrooms. *Foundation for Childhood Development.* Report retrieved on November 28, 2011 from http://www.childcare research.org/childcare/resources/12893.

Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics, 32*(4), 341.

Huang, B., Sivaganesan, S., Succop, P., & Goodman, E. (2004). Statistical assessment of mediational effects for logistic mediational models. *Statistics in Medicine, 23*(17), 2713–2728. doi:10.1002/sim.1847.

Jennings, P. A., & Greenberg, M. T. (2009). The prosocial classroom: Teacher social and emotional competence in relation to student and classroom outcomes. *Review of Educational Research, 79*(1), 491–525. doi:10.3102/0034654308325693.

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods, 11*(4), 815–852. doi:10.1177/1094428106296642.

Malmberg, L. E., & Hagger, H. (2009). Changes in student teachers' agency beliefs during a teacher education year, and relationships with observed classroom quality, and day-to-day experiences. *British Journal of Educational Psychology, 79*(4), 677–694.

Marinell, W., & Pallas, A. (2011). *Teacher turnover in New York City public middle schools.* Paper presented at the Annual Meeting of the Association for Public Policy Analysis and Management, Washington, DC.

Maurer, M., & Brackett, M. A. (2004). *Emotional literacy in the middle school: A 6-step program to promote social, emotional, & academic learning.* Port Chester, NY: National Professional Resources.

Muthen, B., & Muthen, L. (2010). *Mplus user's guide.* Los Angeles, CA: Muthen and Muthen.

National Research Council & Institute of Medicine. (2002). *Community programs to promote youth development.* Washington, DC: National Academy Press.

NICHD Early Child Care Research Network. (2003). Social functioning in first grade: Associations with earlier home and child care predictors and with current classroom experiences. *Child Development, 74*, 1639–1662.

Pianta, R. C., & Allen, J. P. (2008). Building capacity for positive youth development in secondary school classrooms: Changing teachers interactions with students. *Toward Positive Youth Development, 1*(9), 21–40.

Pianta, R. C., Belsky, J., Houts, R., & Morrison, F. (2007). Opportunities to learn in America's elementary classrooms. *Science, 315*(5820), 1795–1796. doi:10.1126/science.1139719.

Pianta, R. C., La Paro, K., & Hamre, B. (2008). *Classroom assessment scoring system: K-3.* Baltimore, MD: Brookes Publishing.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Raudenbush, S. W., Martinez, A., & Spybrook, J. (2007). Strategies for improving precision in group-randomized experiments. *Educational Evaluation and Policy Analysis, 29*(1), 5–29.

Raver, C. C., Jones, S. M., Li-Grining, C. P., Metzger, M., Champion, K. M., & Sardin, L. (2008). Improving preschool classroom

processes: Preliminary findings from a randomized trial implemented in head start settings. *Early Childhood Research Quarterly, 23*(1), 10–26. doi:10.1016/j.ecresq.2007.09.001.

Reyes, M. R., Brackett, M. A., Rivers, S. E., Elbertson, N. A., & Salovey, P. (2012). The interaction effects of program training, dosage, and implementation quality on targeted student outcomes for The RULER Approach to social and emotional learning. *School Psychology Review, 41*(1), 82–99.

Rimm-Kaufman, S. E., Fan, X., Chiu, Y.-J., & You, W. (2007). The contribution of the responsive classroom approach on children's academic achievement: Results from a three year longitudinal study. *Journal of School Psychology, 45*(4), 401–421.

Rivers, S. E., & Brackett, M. A. (2011). Achieving standards in the English language arts (and more) using The RULER Approach to social and emotional learning. *Reading & Writing Quarterly, 27*(1/2), 75–100. doi:10.1080/10573569.2011.532715.

Rivers, S. E., Brackett, M. A., Reyes, M. R., Elbertson, N. A., & Salovey, P. (2013). Improving the social and emotional climate of classrooms: A clustered randomized controlled trial testing The RULER Approach. *Prevention Science, 14*(1), 77–87.

Rivers, S. E., Brackett, M. A., Reyes, M. R., Mayer, J. D., Caruso, D. R., & Salovey, P. (2012). Measuring emotional intelligence in early adolescence with the MSCEIT-YV: Psychometric properties and relationship with academic performance and psychosocial functioning. *Journal of Psychoeducational Assessment, 30*(4), 344–366.

Rutter, M., & Maughan, B. (2002). School effectiveness findings 1979–2002. *Journal of School Psychology, 40*(6), 451–475.

Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, Cognition and Personality, 9*(3), 185–211. doi:10.2190/DUGG-P24E-52WK-6CDG.

Seidman, E. (2012). An emerging action science of social settings. *American Journal of Community Psychology, 50*(1–2), 1–16. doi:10.1007/s10464-011-9469-3.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7*(4), 422–445. doi:10.1037/1082-989X.7.4.422.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420–428. doi:10.1037/0033-2909.86.2.420.

Solomon, D., Watson, M., Battistich, V., & Schaps, E. (1996). Creating classrooms that students experience as communities. *American Journal of Community Psychology, 24*(6), 719–748. doi:10.1007/bf02511032.

Steiner, P. M., Cook, T. D., Shadish, W. R., & Clark, M. (2010). The importance of covariate selection in controlling for selection bias in observational studies. *Psychological Methods, 15*(3), 250–267. doi:10.1037/a0018719.

Sutton, R. E., & Wheatley, K. F. (2003). Teachers' emotions and teaching: A review of the literature and directions for future research. *Educational Psychology Review, 15*(4), 327–358.

Wentzel, K. R. (2002). Are effective teachers like good parents? Teaching styles and student adjustment in ealry adolescence. *Child Development, 73*, 287–301.