THE EFFECT OF PROGRAM IMPLEMENTATION QUALITY ON THE SOCIAL-

EMOTIONAL LEARNING AND NEGATIVE MENTAL HEALTH OF URBAN

MIDDLE SCHOOL STUDENTS IN THE MOSAIC INTERVENTION

by

ARIELLE CLAIRE VANPEE LINSKY

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Psychology

Written under the direction of

Maurice J. Elias, Ph.D.

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

OCTOBER, 2020

ABSTRACT OF THE DISSERTATION

The Effect of Program Implementation Quality on the Social-Emotional Learning and

Negative Mental Health of Urban Middle School Students in the MOSAIC Intervention

By ARIELLE CLAIRE VANPEE LINSKY

Dissertation Director:

Maurice J. Elias, Ph.D.

Social-emotional and character development (SECD) programs promote the

development of youth's social-emotional skills (e.g., empathy) and cultivation of pro-

social character virtues (e.g., gratitude). When provided universally within the school

context, these interventions have been shown to increase students' social-emotional

learning (SEL) and decrease emotional distress. Higher levels of program implementation

have been associated with more positive student outcomes. The acquisition of SEL has

been hypothesized as a pathway between intervention and mental health outcomes.

However, the majority of research has been conducted in schools serving predominantly

middle-class and White students.

The current study sought to address this gap by evaluating the association of

program implementation quality and student outcomes in an SECD intervention

("MOSAIC") in three urban middle schools. Participants were 6-8th grade students,

predominantly students of color from low-income households, who attended an

intervention school for two years ($n = 308$). Utilizing multilevel modeling, program

implementation quality was tested as a predictor of teacher-rated SEL and self-reported

negative mental health in Year 1 and Year 2 of the MOSAIC intervention. Teacher-rated

SEL was also tested as a mediator of program implementation quality and student negative mental health.

Program implementation quality was found to significantly predict student negative mental health in Year 2, such that students with stronger implementation quality presented with less mental health problems. Program implementation quality did not predict SEL in Year 1 or 2. SEL was not a significant mediator between program implementation quality and negative mental health. However, post-hoc analysis revealed that Fall SEL was a significant moderator of program implementation quality in predicting student negative mental health in Year 2. Students with higher SEL at the start of the year appeared to benefit more from high quality program implementation, resulting in fewer mental health problems at the end of the year.

These findings suggest that in the urban middle school setting, high quality implementation of SECD interventions is positively associated with student mental health outcomes. Results also illuminate the complexity of implementation and measurement in this setting and the need for a context-specific approach for future research in this area.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

The Effect of Program Implementation Quality on the Social-Emotional Learning and Negative

Mental Health of Urban Middle School Students in the MOSAIC Intervention

In recent years, many powerful entities in health and education have advocated for the

promotion of social-emotional and character development (SECD) interventions in schools (e.g.,

AEI/Brookings Working Group on Poverty and Opportunity, 2015; Domitrovich, Syvertsen, &

Calin, 2017; Jones & Kahn, 2017). Additionally, many state departments of education across the

United States have implemented, or are developing, social-emotional learning standards

(Dusenbury et al., 2015). SECD interventions are school programs that promote positive youth

development by helping students to acquire social-emotional competencies, including problem-

solving, emotion regulation, self/social-awareness, and responsible decision making, as well

developing character virtues aligned with prosocial values, such as generosity and gratitude

(Elias, 2014)[1].

These reports and policy movements have been influenced, in large part, by three meta-

analyses of SECD school interventions that reported significant student gains in social-emotional

learning (SEL), attitudes, positive social behaviors, and significant decreases in emotional and

behavioral problems (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Sklad,

Diekstra, Ritter, Ben, & Gravesteijn, 2012; Taylor, Oberle, Durlak, & Weissberg, 2017). With

---

[1] It is important to note that these interventions are sometimes identified by other names (social-emotional learning (SEL), character education, and risk prevention, to name a few). However, there is a growing consensus in the field that "social and emotional development is multi-faceted and is integral to academics," and that these approaches may differ by name but not by core essence (Jones & Kahn, 2017, p. 5). For the purposes of this paper they will be referred to as SECD interventions, except when referring to a specific program name or in a quotation, which may use a different term.

strong empirical grounding of the positive impact of SECD in schools on student competencies provided by these meta-analyses, researchers are now tasked with investigating: 1. *What* program components are most effective for *who* and in *which* settings? 2. What are effective methods for wide-spread dissemination, implementation, and sustainability (Mahoney, Durlak, & Weissberg, 2018)?

In particular, many reports place an emphasis on the need for SECD interventions in low-resource schools, often serving racial and ethnic minority youth (e.g., Jones & Kahn, 2017; Simmons, Brackett, & Adler, 2018). For example, the AEI/Brookings Working Group on Poverty and Opportunity (2015) stated:

> "What we need now are policies to scale up high-quality, evidence-based SEL programs and to make them a fundamental part of the education of all kids, but especially children who need it the most—low-income children who will fall further behind without it." (p. 60).

However, to date, the majority of studies of SECD school interventions have been conducted in middle class, majority White, U.S. schools, and very few have evaluated the impact of diversity factors such as race, ethnicity, socio-economic status, and sexual/gender identity on program outcomes (Rowe & Trickett, 2017). Further, we still have much to learn about how program components, as well as quality of implementation, relate to SECD program outcomes. In an analysis of 500 studies from 1976-2006, program quality was rarely included and when it was, it was usually not systematic (Durlak & DuPre, 2008).

In this next phase of intervention implementation, refinement, and dissemination, it is essential that interventions consider issues of cultural diversity, equity, and social action and power (Hoffman, 2009; Romasz, Kantor, & Elias, 2004). SECD in schools has the opportunity to

help all students reach their full potential and effectively contribute to an increasingly intercultural society (Gorski & Covert, 2000). The proposed project aims to evaluate a social-emotional and character development program implemented at the universal level in four urban middle schools in central New Jersey. Called, 'MOSAIC' (Mastering our Skills and Inspiring Character), the intervention builds on current theory, research, and practice that is based largely, as noted earlier, in White middle-class populations, and adapts it to a low-resource, urban middle school context. Using multilevel modeling to account for the clustering of students within classrooms, this study aims to understand the relationship between degree of program implementation quality of the MOSAIC intervention and student social-emotional learning and negative mental health, occurring from Fall to Spring of the first and second years of the intervention. Additionally, this study evaluates social-emotional learning as a mediator of the relationship between program implementation quality and negative mental health.

SECD School-Wide Interventions

The term 'social-emotional and character development' (SECD) calls for the integration of the fields of Social-Emotional Learning (SEL) and Character Education (CE) so students learn to "not only know the right ways to behave, but also to possess and use the skills to enact desired behaviors effectively" (Elias, 2014. p. 37). SECD also explicitly recognizes that student learning is embedded in context, so that an excellent curriculum—in any content area—delivered in an environment where relevance, use, or importance are cast into doubt or delivered poorly, will have no lasting impact. Accordingly, along with a coordinated developmental sequence that progresses with complexity over time, programs must combine explicit social-emotional skill instruction with value/virtue definition along with efforts to promote positive school climate to create a message of reinforcement and support throughout the school day (Elias, 2014). SECD,

in the context of systematic oppression, trauma, and inequity that characterizes many urban, minority environments, can help students increase attributes (e.g., compassionate forgiveness, optimistic future-mindedness) that have been shown to be effective in promoting flourishing and positive development, even in contexts of adversity (see: Hatchimonji, Linsky, & Elias, 2017 for further elaboration).

Social-Emotional Learning (SEL)

Across SECD school intervention results, improvement in social-emotional knowledge and ability has been the most robust and consistent finding. These competencies include:

(1) Self-Awareness: the ability to effectively label one's emotions and values, and assess strengths and weaknesses;

(2) Self-Management: the ability to regulate one's emotions, thoughts and behaviors across a range of situations and to employ this capacity toward coping with stress, working toward goals, and managing impulses;

(3) Social Awareness: the ability to take the perspective of others with different backgrounds, understand social and ethical behavior norms, and identify resources and supports;

(4) Relationship Skills: the ability to effectively communicate, develop healthy and positive relationships, and resolve conflict with others; and

(5) Responsible Decision-Making: the ability to apply problem solving techniques to make constructive choices that take ethical standards, social norms, and safety into consideration (Durlak et al., 2011; Ogden et al., 2016).

In Durlak's meta-analysis of 211 SECD universal school interventions (31% of studies included middle schoolers), 68 studies measured improvement in SEL post intervention. This outcome

had the largest effect size (0.57) of all outcomes measured. At follow-up, effect size remained significant ($ES = 0.26$, $n = 21$) as were mean effects (0.24) of increased positive social behavior (e.g. getting along with others in daily behavior) across the 86 studies that included this outcome measure (Durlak et al., 2011). Interestingly, this effect size remained significant and increased to 0.26 when the intervention was led by a classroom teacher (59 studies analyzed) and dropped to a nonsignificant 0.23 when led by non-school personnel, such as researchers and graduate students (11 studies analyzed). Effect size at follow up remained significant ($ES = 0.17$, $n = 12$). A significant mean effect size of 0.7 was also found amongst the 31 studies in the Sklad and colleagues (2012) meta-analysis that assessed for social-emotional skills. This indicates that an average participant in a program reviewed had better skills than 76% of students who did not participate in the program. Effect size remained significant for the 15 studies that reported on it. However, at follow-up, effect size was small (0.07). Similarly, authors identified six studies assessing for pro-social behavior as an immediate outcome and seven studies that assessed for such behavior at follow-up. Amongst these studies, the immediate outcome yielded a significant effect size of 0.39 and follow-up yielded a significant effect size of 0.12 (Sklad et al., 2012).

In addition to meta-analysis results, it is helpful to examine findings of individual interventions. In a 2010 review of the 15 separate studies conducted by the Oregon Resiliency Project, the creators of the Strong Kids programs reported that, in all studies, participants demonstrated a significant gain in social-emotional knowledge, with large effect sizes (near or above 1.0); (Merrell, 2010). In several studies of adaptations involving Strong Kids intervention published after Merrell's 2010 review, findings continued to indicate increases in knowledge of social-emotional skills and self-reported resilience.

A Spanish-language adapted version, 'Jovenes Fuertes,' was implemented in five middle and two high schools in southern California. With randomly assigned Spanish-dominant ELL classrooms ($n = 102$), students in the intervention arm showed significantly higher self-reported knowledge of social-emotional skills and levels of resiliency (rated on the Behavior-Emotional Rating Scale-2); (Castro-Olivio, 2014). When a version of this program designed for elementary school students, Strong Kids, was implemented with predominantly White third and fourth graders in the Pacific Northeast ($n = 106$), the treatment group scored higher on measures of social-emotional skill knowledge, as well as teacher and self-rated use of social skills, at both the post-test and two-month follow-up, when compared to waitlist controls (Harlacher & Merrell, 2010).

Indicators of Negative Mental Health

One in five adolescents in the United States is diagnosed with a mental illness, most commonly a mood or anxiety disorder, many of which will persist into adulthood (Lee et al., 2014). In recent years, well-respected organizations have called attention to this issue. The Institute of Medicine's 2009 report called for the federal government to prioritize prevention of youth mental, emotional, and behavioral disorders, suggesting a budget of $247 billion annually (O'Connell, Boat, & Warner, 2009). One year later, results from the National Comorbidity Survey—Adolescent Supplement (NCS-AS), a national survey of over 10,000 13-17 year olds conducted by Harvard Medical School (see: https://www.hcp.med.harvard.edu/ncs/index.php), found that 48.3% of Latino, 46.8% of Black, and 41.9% of non-Latino White participants reported experiencing a mental health disorder before the age of 18 (Merikangas et al., 2010).

Although indicators of negative mental health are less frequently assessed in school-based SECD studies, existing research supports SECD interventions as efficacious in decreasing

indicators of negative mental health. In Durlak et al.'s (2011) meta-analysis, 49 studies assessed for emotional distress, and a significant decrease in symptoms (*ES* = .24) was found across studies. Of these studies, 11 included a follow-up of at least 6 months post-intervention, and the effect size of emotional distress decrease (0.15), though smaller, remained significant. Only 13 of the 75 studies evaluated in the Sklad et al. (2012) meta-analysis assessed for internalizing symptoms as an immediate outcome, and 11 evaluated this outcome at follow-up. Effect sizes for decrease in symptoms were significant for both immediate (0.19) and follow-up (0.10) outcomes. In a recent randomized control trial of a CASEL-select2 SECD intervention, Positive Action (PA), students in intervention schools reported significantly lower anxiety and depression symptoms at endpoint than those in control schools (Lewis et al., 2013). Taylor et al. (2017) found that across 35 studies, students who participated in SECD programs reported significantly lower emotional distress (including symptoms of anxiety and depression) at follow-up than peers in control groups (ES = 0.16, mean follow-up = 88 weeks, improvement index = 5.64%). Taken together, research supports that universal, school-based interventions are associated with decreases in indicators of negative mental health. However, it is worth noting that very few if any of these studies were carried out in urban, minority, low-SES contexts.

Relationship of Social-Emotional Competencies and Negative Mental Health

Several recent publications have begun to illuminate the pathway from social-emotional competencies and mental health. In the recent Randomized Control Trial (RCT), conducted in urban K-8 schools serving predominantly low-income students in Chicago, Illinois, evaluating the elementary school-based *PA* SECD intervention, socio-emotional skills and behaviors

---

2 The Collaborative for Academic, Social and Emotional Learning (CASEL) Guide provides educators with recommendations for SEL/SECD programs that are considered evidence-based. More information can be found at https://casel.org/guide/criteria/

significantly mediated the increase in positive affect and life satisfaction, along with the decrease in anxiety and depression at study end (Lewis et al., 2013). In longitudinal evaluations, strong associations have been shown between the development of specific social-emotional competencies in adolescence and positive adjustment outcomes (e.g., mental health, academic success, and healthy relationships) (Guerra & Bradshaw, 2008). Additionally, the absence of social-emotional competencies in adolescence is associated longitudinally with negative outcomes (e.g., higher rates of criminal, substance abuse, and mental health issues) (Modecki, Zimmer-Gembeck, & Guerra, 2017).

The 'Social and Emotional Learning Framework for Positive Youth Development' (Taylor et al., 2017) posits a similar pathway: that school-based, social-emotional learning interventions cultivate skills within participating youth which promote the development of positive indicators of well-being (e.g., academic success and positive social behavior) and deter the development of negative indicators of well-being (e.g., emotional distress, conduct problems, and drug use). Enhanced social-emotional skills, though not attitudes, significantly predicted improved indicators of well-being at follow-up across the 42 studies included in this analysis (Taylor et al., 2017). Taken together, an evidence base is beginning to develop demonstrating the mediating role of acquisition of social-emotional competencies between SECD intervention and mental health outcomes. However, as will be discussed next, this knowledge base has not been established across contexts.

Program Implementation

Program implementation, or the extent to which a program is conducted in the way it was intended, is often problematic in school and community settings (Gottfredson & Gottfredson, 2003). A review of school-based drug abuse prevention programs over the last quarter of the 20th

century found that poor implementation is associated with weaker program effectiveness; that the majority of teachers do not teach the full curriculum; and that the amount they teach decreases over time (Dusenbury, Brannigan, Falco, & Hansen, 2003).

This is not surprising given the plethora of challenges facing researchers and practitioners when aiming to implement a program in the school setting. For SECD interventions specifically, many teachers—the most effective implementers of SECD (Durlak et al., 2011)—are skeptical of SECD program effects. Among other concerns, they indicated that SECD should be the job of the school psychologist or counselor rather than the teacher; that SECD activities will take time away from test preparation and academics; and that other prevention programs (e.g., smoking prevention) meet SECD goals, among other concerns (Elias, Bruene-Butler, Blum, & Schuyler, 2000). Teachers' attitudes and beliefs about the intervention can impact their effectiveness as implementers. Teachers can feel particular resistance to a program if not involved in the decision to adopt the program (Greenberg, Domitrovich, Graczyk, & Zins, 2005). Teachers also have varied levels of comfort with teaching SECD curricula. In a survey of 263 kindergarten-to-eighth grade teachers from two states regarding knowledge, perceptions, and practices of SEL, only 15.9% indicated that they were satisfied with their current knowledge of SEL and only 22% reported that they were confident in their ability to implement SEL (Buchanan, Gueldner, Tran, & Merrell, 2009).

SECD implementation is certainly challenging. But research demonstrates that quality implementation is important to program outcomes. In a study of almost 5,000 students receiving the KidsMatter SECD intervention, investigators found that students' social-emotional competencies only significantly improved in average and high implementing schools, but not in low implementing schools (Askell-Williams, Dix, Lawson, & Slee, 2013). Another study of the

Promoting Alternative Thinking Strategies (PATHS) program in urban district in Pennsylvania serving predominantly White, low-SES students found that teacher burnout was negatively associated with implementation dosage and teacher efficacy was positively associated with implementation dosage. Further, teacher perception of support from school administration was positively associated with implementation quality (Ransford et al., 2009). Only 42% of studies included in Durlak et al.'s meta-analysis reported on implementation factors. Of these studies, those that reported any implementation problems, such as failure on the part of the staff to conduct intervention components, or unforeseen circumstances that impacted the intervention execution, found that implementation moderated outcomes. These interventions showed significant results in only two of the six evaluated areas – attitudes and conduct problems. Thus, programs reporting implementation problems did not support significant outcomes in social-emotional competency development or reduction emotional distress (Durlak et al., 2011; Gager & Elias, 1997).

Implementation contains multiple dimensions. Although specific designations of dimensions may vary, five overarching aspects are often cited: adherence/fidelity, dosage, quality, participant responsiveness, and program differentiation (Dane & Schenider, 1998; Durlak & Dupree, 2008; Schult, Easton, & Parker, 2009). For SECD interventions, the dimension of program implementation quality appears to be particularly important in impacting program outcomes.

Program implementation quality refers to *how* an intervention is executed and includes both the delivery and attitudes surrounding execution. Due to the difficulty of assessing this complex construct, it is rarely looked at. In a review of 500 studies from 1976-2006, the construct was rarely included. When it was, it was usually not systematic (Durlak & DuPre,

2008). However, in a recent analysis of the RULER Program, implementation quality was assessed as a predictor of student social competence, problem solving, and emotional literacy. This study evaluated 812 sixth-grade students in 28 elementary schools within an urban Catholic school district in the northeastern United States. About 23% of students in study schools qualified for free or reduced lunch and students in study sample were 27% White/Caucasian, 30.4% Black/African American, 22% Hispanic, 7.5% Asian/Pacific Islander, and 3.7% multiracial. Implementation quality was measured by ratings by RULER coaches at the beginning and end of the school year indicating the extent to which teachers (1) demonstrated buy-in/open attitude to the program and (2) delivered RULER with high quality. Although these measures were submitted at only two time points, the RULER coaches had met with teachers multiple times (up to 5) over the course of the year. Results indicated that high implementers with more dosage and training had students with higher outcomes of social competence, problem solving and emotional literacy while moderate implementers with more training had students with higher emotional literacy and more dosage had students with higher social competence and social problem-solving (Reyes, Brackett, Rivers, Elbertson, & Salovey, 2012).

One reason implementation quality is so difficult to assess is because accurate assessment requires comprehensive understanding of both the program and the context (Waltz, Addis, Koerner, & Jacobson, 1993). Although the gold-standard for measuring implementation quality for SECD programs has not been established, coach or consultant evaluations are among the typical methods (along with teacher self-ratings and direct behavior observations) (Durlak, 2016). In the evaluation of the RULER program, implementation quality was measured by coaches who reportedly met with teachers multiple time points (on average 4) over the course of the year, were extensively trained in the RULER intervention themselves, and were supervised

through regular meetings and record review by a senior RULER trainer (Reyes et al., 2013). In addition to the expertise required to effectively rate program quality, it is important that raters observe and meet with implementers at multiple time points in order to incorporate information across the entirety of the intervention are likely to be more accurate (Durlak, 2016).

Cultural and Ecological Considerations

The relationship of race, ethnicity, and socio-economic status with social-emotional competencies and mental health is not entirely straightforward. The recently published 'Conceptual Model for Child Mental Health and Mental Health Service Disparities' (Alegría, Green, McLaughlin, & Loder, 2015) aims to effectively integrate research and theory. This model posits that four mechanisms—socio-economic status, childhood adversities, family structure across development, and neighborhood-level factors—influence the trajectory of mental health for minority youth in the U.S. These mechanisms each affect minority individuals disproportionately. People of color are more likely to live in poverty, experience childhood adversity and trauma, have family structures associated with negative outcomes (e.g., single-parent households), and live in neighborhoods struggling with risk factors such as violence and educational inequities. Thus, each mechanism must be considered as part of an interconnected network interacting in a bi-directional relationship with minority youth, and contributing to mental health disorder in complex ways (Alegría et al., 2015). For this reason, prevention efforts to address this inequity must focus not only on individuals, but also on the contexts in which they reside. The strength/intensity, or dosage, of the intervention is another associated consideration in disadvantaged settings.

Economically disadvantaged youth are more likely to go to schools with environmental and relational factors that can hinder their developmental trajectory, including punitive

environments, less robust academics, and higher numbers of underprepared teachers, as well as

high turnover among staff (Osher et al., 2016). Youth exposed to continuous stress and trauma

are particularly vulnerable to social-emotional skill deficits and thus may experience the same

environment in a different way from their peers (Jones & Kahn, 2017). SECD programs have the

opportunity to influence these environments and improve student outcomes, but they also must

adapt to the conditions experienced by both students and staff.

The majority of SECD studies to date have not included samples comprised of primarily

students of color; in the Durlak et al. (2011) meta-analysis, of papers that included follow-up

data ($N = 117$) only 64% even reported race/ethnicity. Of these 75 studies, 34 included a

primarily White sample, 17 primarily Black, only 1 predominantly Latino, 1 primarily Native

American/Alaska Native, and no primarily Asian-American samples. SES was reported in 45%

of studies and was described in inconsistent ways (Rowe & Trickett, 2017). Although current

research provides evidence that SECD interventions can be effective across demographic groups,

it is difficult to draw firm conclusions based on the small number of studies of minority

populations included, particularly of Latinx students (Sklad et al., 2012; Taylor et al., 2017).

Operating within the scarcity of empirical evidence, scientists working in urban middle

schools serving predominantly students of color from low-income households are guided largely

by theory. Victor Frankl illuminated the power of positive purpose to survival of the spirit and

body in the Nazi concentration camps. The analogy to the youth developing with the challenging

personal, school, and community contexts of the low-resource urban setting is not a large stretch.

SECD programs that help students to cultivate character virtues and a purpose relevant to the

challenges they face (e.g., forgiveness and future-mindedness) in conjunction with the social-

emotional competencies to enact those virtues and purpose, are likely to be most effective (See Hatchimonji, Linsky, & Elias, 2017 for further elaboration on this theory).

The Current Study

The current study was part of a larger three-year grant, funded by the John Templeton Foundation, entitled, *Enhancing Student Purpose with the Middle School Ambassador Collaborative Action-Research Study (ID #56203).* The larger study included the implementation of an SECD intervention in six middle schools in a large, urban district in Central New Jersey. The intervention, called MOSAIC ("Mastering Our Skills and Inspiring Character") was based on a pilot over the 2012-2014 school years in grades 6-8 another Central New Jersey urban middle school which serves predominantly Hispanic students from low-income families. The intervention used a collaborative action-research approach to develop a program that fit the setting, based on the lessons learned in the pilot. MOSAIC consists of 15-minute daily lessons taught during advisory periods or social studies classes to 6th, 7th, and 8th grade students, with curriculum developmentally differentiated by grade level. Classes included virtue definition and cultivation, skill practice, and a social action lessons series (called "Students Taking Action Together" or "STAT") for one week per month and to connect with the greater school community to enact social action projects (called "Positive Purpose Projects").

MOSAIC was designed specifically to adapt evidence-based SECD interventions to the context of an urban middle school setting. The MOSAIC advisory lessons taught students to cultivate their noble purpose through explicit instruction on five supporting virtues (constructive creativity, helpful generosity, optimistic future-mindedness, responsible diligence, and compassionate forgiveness/gratitude) and four focal SEL skills (emotional regulation, communication, empathy, and social problem solving), organized around a framework of themes

that connect the focal virtue and skills of the month (see Appendix A: MOSAIC Virtues & Skills for breakdown of skills, virtues, and themes by calendar month; see also Hatchimonji, Linsky, & Elias, 2017 for further discussion on theory of the cultivation of noble purpose through MOSAIC).

MOSAIC lessons were differentiated by grade-level, with 6th graders focusing on applying the skills and virtues to themselves, 7th graders to their school and immediate community, and 8th graders to the greater world. Additionally, monthly 'Throughline Sheets,' aimed at prompting students and teachers alike to practice MOSAIC lessons throughout the day, were posted in every classroom in participating schools. Advisory lessons used pedagogical methods, such as narrative examples and values clarification exercises, to introduce virtues. Instructional techniques such as debates, role-plays, games, and discussion were used to facilitate skill and virtue acquisition. Reflection and feedback lessons were facilitated as part of the curriculum every other month from teachers and students to encourage active participation in the development of program (see Hatchimonji, Linsky, DeMarchena, Kim, Nayman, & Elias, 2017 for further explanation of the MOSAIC participatory feedback process). The pedagogical assumption was that students had not experienced systematic SECD instruction in their elementary schools, so that MOSAIC would be their first exposure to this kind of curriculum.

The social-action lessons were referred to as Students Taking Action Together (STAT) and used a service-learning framework modeled after the PARD-C method (preparation, action, reflection, demonstration, and celebration) developed by Kaye (1997). In addition to leading STAT classes in their MOSAIC advisories, student ambassadors also participated in regular meetings to work on larger-scale social-action projects called Positive Purpose Projects, such as cleaning and decorating the school blacktop, leading a diversity-oriented field day, or facilitating

a school assembly to address bullying. At the end of the school year, ambassadors from all six schools attended a showcase to demonstrate their Positive Purpose Projects and celebrate their efforts with a larger community (see Linsky, Hatchimonji, Kruzik, Kifer, Franza, McClain, Nayman, & Elias, in press, and https://www.secdlab.org/ambassadors for more information about the social action components of MOSAIC).

The implementation of the MOSAIC intervention was supported at several levels in each school. Each school was assigned a MOSAIC consultant who worked with that school for the entire school year. MOSAIC consultants were graduate student clinicians with training and expertise in school psychology and SECD. MOSAIC consultants visited their assigned school weekly to meet with key stakeholders and observe and support MOSAIC implementation. MOSAIC consultants met weekly as a group with the Principal Investigator for supervisory purposes and information exchange. Consultants also took detailed process notes after each school visit was completed and circulated these notes to the entire team for questions and comments (see Appendix C for the MOSAIC Consultant School Visit Note Checklist). School administrators met typically monthly with the MOSAIC Principal Investigator, as well as with MOSAIC school consultants, to discuss the big picture roll-out of the intervention.

Each school designated an internal MOSAIC point-person (usually a teacher or counselor) who served as the liaison between the research team and the schools. Some schools formed a larger committee that met regularly to support the implementation of MOSAIC in the school. The MOSAIC Research Team (including consultants) provided training to teachers and staff implementing MOSAIC lessons whenever possible throughout the year. The team also reviewed bi-monthly curriculum feedback immediately in order to respond with program adaptations in real-time.

Research Questions and Hypotheses

The primary aim of this research project is to test the impact of program implementation quality of the MOSAIC Program on the development of middle schoolers' teacher-rated social-emotional learning (SEL) and self-reported negative mental health. The study will look at the same sample of students in both years of the intervention, utilizing separate models for each year due to changes in classroom cluster between Year 1 and Year 2. The project aims to evaluate Spring outcomes, controlling for Fall baseline, for student participants in the Year 1 and Year 2 of MOSAIC Program implementation. Specifically, the project aims to answer the following two research questions.

Research question 1. Does stronger MOSAIC program implementation quality over the course of each of two intervention years (Year 1 and Year 2) predict an increase in teacher-rated social-emotional learning from Fall to Spring of each intervention year (Fall Year 1 – Spring Year 1 and Fall Year 2 – Spring Year 2)?

Hypothesis 1a. Stronger MOSAIC program implementation quality in Year 1 will predict greater teacher-rated SEL in Spring Year 1 when controlling for teacher-rated SEL in Fall Year 1.

Hypothesis 1b. Stronger MOSAIC program implementation quality in Year 2 will predict greater teacher-rated SEL in Spring Year 2 when controlling for teacher-rated SEL in Fall Year 2.

Research question 2. Does stronger MOSAIC program implementation quality over the course of each of two intervention years (Year 1 and Year 2) predict a decrease in self-reported negative mental health from Fall to Spring of each intervention year (Fall Year 1 – Spring Year 1 and Fall Year 2 – Spring Year 2)?

Hypothesis 2a. Stronger MOSAIC program implementation quality in Year 1 will

predict lower self-reported negative mental health in Spring Year 1 when

controlling for negative mental health in Fall Year 1.

Hypothesis 2b. Stronger MOSAIC program implementation quality in Year 2 will

predict lower self-reported negative mental health in Spring Year 2 when

controlling for self-reported negative mental health in Fall Year 2.

Research Question 3: Is the acquisition of teacher-rated SEL a pathway between

program implementation quality and self-reported negative mental health in each

intervention year (Year 1 and Year 2)?

Hypothesis 3a. The residual change score of Fall Year 1 to Spring Year 1 teacher-

rated SEL will be a significant pathway between MOSAIC program

implementation quality Year 1 and self-reported negative mental health in Spring

Year 1.

Hypothesis 3b. The residual change score of Fall Year 2 to Spring Year 2 teacher-

rated SEL will be a significant pathway between MOSAIC program

implementation quality Year 2 and self-reported negative mental health in Spring

Year 2.

Methods

Participants

This study was part of a larger study that took place in six urban middle schools in

Central New Jersey during the academic years 2015-16 and 2016-2017. Four of these six schools

were selected for the current study to provide a balanced representation of school makeup (e.g.,

6-8 or k-8) and school-level program implementation quality, based on observations of the

consulting team and specific factors to be discussed subsequently. The four schools included two standalone 6-8 grade middle schools, one with higher school-level program implementation quality (School D, *n* = 144) and one with lower school-level program implementation quality (School A, *n* = 521), and two K-8 elementary schools, one with higher school-level program implementation quality (School B, *n* = 187) and one with lower school-level program implementation quality (School C, *n* = 246). The study evaluated students who were in intervention schools during *both* the 2015-16 and 2016-17 academic years. Schools had no significant differences in gender or grade. Although the chi square test was significant for differences in household income (measured by eligibility for free or reduced lunch), all four schools were made of majority students from low-income households (73.7% of sample). With regards to racial/ethnic make-up, three schools had a majority of students identified as Hispanic (44.4% of sample) and the next largest group were students identified as Black (26.7% of sample). In School D, 54% of students were identified as Black and 28.5% as Hispanic. Additionally, School D had a larger percentage (23.6%) of students receiving special education services than other schools. See Table 1 for complete demographic information relating to the original sample.

Procedures

In the Fall Year 1 and Year 2, students were consented to participate in the study through a passive consent process approved by the school board and the research institution's IRB. Student self-assessment batteries were given twice (Fall and Spring) during each intervention year. In Fall Year 1, surveys were administered electronically through the web-based system, Qualtrics. In response to school feedback, subsequent student assessments (Spring Year 1, Fall Year 2, Spring Year 2) included the following changes: 1. Item deletion to decrease assessment

length, create a more parsimonious assessment based on factor analysis, and respond to school concerns about certain items. 2. Change from web-based survey to paper-pencil scantron survey. MOSAIC teachers completed the Devereux Student Strengths Assessment-mini (DESSA-mini) in Fall and Spring Year 1 and Year 2 using a Qualtrics survey during scheduled prep time as determined by individual school administrators (estimated survey completion time of one minute/student). See Appendix B for data collection table.

Measures

Demographics. Student demographics, including race/ethnicity, grade level, free/reduced lunch status, age, and gender were collected from the district's student database.

Program implementation quality. Program implementation quality was measured by the MOSAIC Consultant Rating for study Year 1 and Year 2. Each MOSAIC school was assigned one MOSAIC Consultant ($n = 3$) for each intervention year. Consultants visited assigned MOSAIC school(s) weekly for 1-4 hours throughout the course of the study (see Current Study Section for further explication of consultant role). Weekly consultant group supervision by the Principal Investigator (in person and review of notes over email) created a normative understanding of MOSAIC program implementation quality across schools. Based on the year's weekly visits and reflection on notes and supervision with MOSAIC teammates, MOSAIC consultants completed a global rating of each MOSAIC teacher's intervention performance in their assigned school. On a Google form, consultants answered the prompt, "How would you rate the performance of the MOSAIC teacher from [SCHOOL NAME]." Response options were: "Teacher performed POORLY as a MOSAIC teacher and did not deliver MOSAIC curriculum", "Teacher most likely delivered MOSAIC curriculum as desired (nothing extremely good or bad)", "Teacher performed as a STAR and went above and beyond MOSAIC

curriculum", or "Do not have a good enough idea about teacher to judge at all." This rating structure is similar to that used in the evaluation of program implementation quality of the RULER intervention discussed earlier (Reyes et al., 2012).

Social-emotional learning (SEL). The Devereux Student Strengths Assessment- Mini, Form A (DESSA-Mini, Naglieri, LeBuffe, & Shapiro, 2009) is a validated universal screening and progress-monitoring tool for kindergarten-8th grade students' social-emotional competence. The DESSA-Mini is an 8-item scale that is based on a 72-item measure with one overall social competency score (Cronbach's alpha = .97 at baseline). Using a representative sample of U.S. youth, the scale has been standardized and converted into T-scores indicating levels of social-emotional competence provided by the scale developers. The DESSA-Mini asks teachers to rate the extent to which they have observed a student exhibiting specific positive behaviors during the past four weeks on a 5-point Likert scale (0: Never, 1: Rarely, 2: Occasionally, 3: Frequently, 4: Very Frequently). The eight items ask about responsibility, nice action for another, speaking positively, attention, group relationships, performing steps in order, showing care on school-work, and following advice of trusted adult (See Appendix D for full measure).

Negative mental health. The Strengths and Difficulties Questionnaire, Self-Report (SDQ; Goodman, 1997) is a 25-item behavioral screening measure developed for 11-17 year olds. It is comprised of five subscales that contribute to a total score: emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems, and pro-social behaviors. Students are asked to respond "based on how things have been for you over the last six months" in accordance with the Likert Scale, "Not True", "Somewhat True", or "Certainly True". See Appendix D for complete measure. As of 2016, the authors of the SDQ had collected a database of over 4,000 published articles using SDQ measures across a variety of samples (see:

http://time point.sdqinfo.com/py/sdqinfo/f0.py). The data has shown to validly discriminate between a community sample and a sample of individuals with indicators of negative mental health (Goodman, Meltzer, & Bailey, 1998). In the baseline (Fall 2015) assessment, students were asked 20 items (all items except for the prosocial scale) ($\alpha = .82$). At the subsequent time points, the scale was shortened to 11 items (Cronbach's alpha at baseline: $\alpha = .75$) (see Procedures section for explanation of SDQ adaptation).

Preliminary Results

Sample Definition

There were 1,098 sixth to eighth grade students in the original sample (e.g., registered for both Year 1 and Year 2 in the four schools in this study according to demographic data given by the district). Students were included in this study if they completed at least one time point of negative mental health assessment (defined as completing > 75% of the SDQ at any one time point between Fall Year 1 and Spring Year 2), had at least one time point of teacher rating of social-emotional learning (defined as completing > 75% of the DESSA at any one time point between Fall Year 1 and Spring Year 2), and were in a classroom with a program implementation quality rating in *both* study years (defined as a MOSAIC Consultant Rating of low, average, or high program implementation quality in Year 1 and Year 2). Students were removed from the study if they did not have any completed time points of the negative mental health assessment ($n = 57$), they did not have any completed time point of the teacher ratings of SEL ($n = 39$), they were missing program implementation rating in Year 1 ($n = 156$[3]) or Year 2 ($n = 4$), or the consultant indicated they "did not have enough information" to rate the program

---

[3] One study school implemented MOSAIC in select social studies classes, thus some students did not receive the intervention, accounting for the majority of missing consultant ratings.

implementation quality in Year 1 ($n = 276$) or Year 2 ($n = 140$). Students were also removed if they provided an invalid response on the SDQ[4] at any time point ($n = 109$). After removing these participants, only 8 participants from School D remained. Because of this small n, these students were also removed from the final study sample. See Appendix E for a consort diagram depicting sample creation. Demographics of the selected analysis sample ($n = 309$) can be found in Table 2.

Students in the selected analysis sample ($n = 309$) were not significantly different than unselected students in the study schools during Year 1 and Year 2 in regard to gender or eligibility for free/reduced lunch. The students selected for the study analysis were less likely to identify as Black and more likely to identify as Hispanic compared to students not selected for the sample ($\chi_2 (6) = 20.03$, $p = .003$). Additionally, students in the analysis sample were less likely (8% compared to 18%) to receive educational accommodations compared to the unselected students ($\chi_2 (1) = 12.78$, p < .001). This difference may be related to (a) the over-representation of students receiving accommodations in School D who were excluded from the sample for reasons stated above, and (b) some students receiving accommodations who were placed in self-contained classrooms where MOSAIC or study assessments were not implemented. With regards to outcome variables (DESSA and SDQ), students in the analysis sample were more likely to have lower SDQ scores in Spring Year 1 ($t = -5.69$, $p < .000$) and Spring Year 2 ($t = -9.50$, $p < .001$) and higher DESSA scores in Fall Year 2 ($t = 2.61$, $p = .009$), and Spring Year 2 ($t = 3.61$, $p < .001$) compared to unselected students in the study schools.

---

[4] If student responded outside of the response range for SDQ measurement scale (e.g. responded "4" or "5" on a 3-point scale) at any time point of data collection, they were removed from the select study sample.

Missingness at item level. Missingness of individual items on the DESSA and SDQ were examined (e.g., for any time point that was >75% completed, but was missing 1-2 items). Item missingness on the DESSA was quite rare, but when missing, the predominant pattern of missingness was to miss one of the eight items (.09% - 0.1% across all four time points). Similarly, the predominant missing item pattern on the SDQ was to miss one item, and this occurred at a slightly more frequent rate than on the DESSA (.02% - 2.5% across all four time points). Missingness at the item level affected a very small portion of the select analysis sample. For any participant with missing items remaining (e.g. SDQ or DESSA waves with 75-99% completion) mean substitution was used (SDQ: .01 - .03% of cases affected at each wave; DESSA: .00 - .02% of cases affected at each wave). This procedure was unlikely to impact study outcomes. Missingness that is less than 5% in large datasets is understood to be minimally impactful on analytic outcomes (Tabachnik & Fidell, 2013, p. 97).

Normality. Examination of histograms, normality probability plots, and de-trended normality probability plots indicated a negative skew on SDQ mean scores at all time points (lower scores indicate less mental health problems), and the Shapiro-Wilk test was significant. Skewness ranged from 0.46 to 0.65. Kurtosis ranged from -0.25 to 0.40 across time points, indicating that at 3 time points (Fall Year 1, Spring Year 1, and Spring Year 2) the distribution was pointy and heavy-tailed, but in Fall Year 2 the distribution was flat and light-tailed. The same series of analyses indicated a positive skew on the DESSA total scores at all time points (positive scores indicate greater presence of SEL skills). Skewness ranged from -0.69 to -0.24. Kurtosis ranged from -0.56 to 1.03 across time points, demonstrating that DESSA scores were flat and light tailed for all time points except in Spring Year 2, when they were pointy and heavy tailed. Interestingly, DESSA scores appeared to cluster around scores that are multiples of 4

(e.g., 8, 16, 24, and 32). This pattern appeared the most clearly in the Spring Year 1, when 96 of the 158 total scores were multiples of four. This pattern suggests that teachers may have had a tendency to respond to each prompt in an all-or-nothing framework, indicating that the student either exhibited the skill (rating a "4") or did not (rating a "0"). Finally, the same set of analyses were conducted to evaluate normality for the MOSAIC Consultant Ratings in Year 1 and Year 2 and for both time points the distribution was pointy and heavy-tailed, with the majority of students in classrooms rated as "average." Due to the overpowered sample size likely contributing to the significant tests of normality (Fields, 2013, pp. 184), normality transformations were not conducted.

   Outliers. Examination of boxplots indicated univariate, high score outliers on SDQ mean scores at all time points and a low score outlier at one wave (Spring Year 2) for DESSA total scores. Each identified outlier case was examined for response validity and three SDQ outlier responses were determined to be invalid due to responder uniform elevated response across all assessment items, including reverse items. These responses were not included in study analyses Due to invalid response, one participant became ineligible for inclusion in analysis study because student did not have a valid SDQ at any time point. This case was deleted from analysis (new select analysis sample $n = 308$). The one DESSA outlier was determined to be invalid due to uniform low response from rater which did not align with past ratings or other study indicators. This response was not included in study analyses.

   Cook's D, and Leverage Values revealed no concerning multivariate outliers for the relationship of Year 1 MOSAIC Consultant Ratings and Fall 2016 DESSA and SDQ scores. With regard to the relationship between Year 2 MOSAIC Consultant Ratings and Spring Year 2 DESSA and SDQ scores, Cook's D was larger than $4/n$, therefore indicating that outliers may be

influential on analysis of the relationship. However, Leverage Values were not concerning for these relationships.

Descriptive Statistics

Mean and standard deviations of SDQ and DESSA at all time points are found in Table 3. A series of t-tests were conducted to assess differences between Fall and Spring of Year 1 and Year 2 for each study variable. No significant differences were found between Fall and Spring of Year 1 for DESSA or SDQ. In Year 2, Spring DESSA scores were significantly higher than Fall DESSA Scores (t = -2.72 (145), p = .007). The difference between SDQ scores between Fall and Spring of Year 2 approached significance, but in an unexpected direction, with scores increasing over time (t = -1.85 (150), p = .067). MOSAIC Consultant Ratings were also compared between Year 1 and Year 2 at the student level and nonsignificant differences were found.

Figures 1-4 show change across levels of program implementation quality from Fall to Spring in Year 1 and Year 2. Figure 1 shows that students with "poor" and "star" MOSAIC Consultant Ratings decreased slightly in self-reported negative mental health between Fall and Spring Year 1, but students with "average" MOSAIC Consultant Ratings increased. Figure 2 shows that students with "poor" and "star" MOSAIC Consultant Ratings increased in teacher-rated SEL from Fall Year 1 to Spring Year 1, but students with "average" ratings decreased slightly. Figure 3 depicts that from Fall Year 2 to Spring Year 2, students with "poor" MOSAIC Consultant Ratings increased in self-reported negative mental health, while those with "average" or "star" ratings stayed relatively stable. Figure 4 shows that students with "star" and "poor" MOSAIC Consultant Ratings increased in teacher-rated SEL from Fall Year 2 to Spring Year 2, and those with "star" ratings had lower scores than counterparts at both Fall and Spring, but a

steeper positive increase in teacher-rated SEL acquisition. Students with "average" MOSAIC Consultant Ratings decreased slightly.

Relationships Between Study Variables

Mean SDQ scores were significantly positively correlated to each other at each time point, as were total DESSA Scores (see Table 4). As shown in Table 4, mean SDQ and total DESSA scores were significantly negatively correlated at nearly all time points (note that lower SDQ scores indicates less negative mental health). Table 5 depicts Spearman's rho correlation coefficients of program implementation quality variables with residual change scores of SDQ and DESSA over Year 1 and Year 2 were assessed, and all were nonsignificant.

A sequence of independent paired-sample t-tests and one-way ANOVAs were conducted to examine the relationship of demographic variables (gender, school, race/ethnicity, free/reduced lunch, and special education accommodations) with SDQ and DESSA Spring Year 2 scores to inform covariate selection. Female-identified students, students enrolled at School A, and White and Asian students were more likely to have higher DESSA scores in the Spring Year 2. Black and Hispanic students were more likely to have higher SDQ scores (negative mental health) in the Spring Year 2. Additionally, an independent t-test on classroom MOSAIC curriculum condition (Ambassador-led or Teacher-Led social action lessons) indicated that students in the Ambassador-led condition were more likely to have higher DESSA scores in Spring Year 2 than students in Teacher-led conditions. Thus, gender, school, race, and MOSAIC curriculum condition were included in multilevel models as covariates. Note that race was transformed to a binary variable (1 = White or Asian, 0 = Black, Hispanic, or other) because the t-tests demonstrated this breakdown into two groups explained the significant findings and because this allowed for more parsimonious interpretation of multilevel model findings.

Time Point Missingness on SDQ and DESSA

Participants also exhibited time point level missingness, when a time point of the SDQ or DESSA was not completed[5] (see Table 3 for percentage of wave missingess by measure). Mean SDQ and total DESSA scores failed to meet Little's test of missing completely at random (MCAR). Missing measures were established to be missing at random (MAR), meaning that the "probability of missingness is unrelated to unobserved concurrent outcomes" (Singer & Willett, 2003, p. 159). MAR is an essential designation in order to utilize established missing data procedures. For the current study, establishing MAR means that participants' missing wave of SDQ or DESSA data should be unrelated to their SDQ or DESSA scores. Because student completion of the SDQ was largely dependent on school staff's administrative capacity, missing SDQ measures are not likely to be reflective of student response. Similarly, missing DESSA measures were largely dependent on administrative support for teachers' completion and not reflective of teachers' perspective on students or of students' SEL competencies.

There are many different views regarding treating missing data in multilevel modeling (e.g., Schafer, 1997; Sterne et al., 2009). Multiple imputation has been cited as a method for decreasing bias and maintaining power compared to a complete case analysis approach (Rubin, 1987). However, many have pointed out that this method has been inadequately evaluated and understood for multilevel models (Schafer, 1997; van Buuren, 2011). Although flat-file multiple imputation, which ignores the hierarchical clusters, has sometimes been discouraged in multi-level modeling, van Buuren (2011) suggests that it is preferable to complete case analysis or

[5] Some administrations of the SDQ had several items completed, but did not meet the >75% completion minimum. These time points were marked as missing, but the participant was retained in the sample (three administrations of the Fall Year 1, two Spring Year 1, and one from Spring Year 2).

single imputation and this is typically a recommended procedure. Further, newer methods aimed at accounting for hierarchical nature of the data (e.g., BLImP and R/JAGs) require advanced statistical ability and are not yet well understood or feasible to conduct for many researchers.

Thus, for this study, multiple imputations for missing time points of SDQ and DESSA scores were conducted using the SPSS 24 multiple imputation feature. This feature was used to create 5 imputed datasets imputing for SDQ and DESSA at all time points (note that van Buuren et. al., 1999 denotes the choice of 5 imputed datasets as "conservative" (p. 686).) Linear regression imputation (as suggested by Rubin (1987)) maximum percentage of missing values of 50%, and maximum number of parameters in imputation modes of 100 were set as specifications. Following the guidelines of van Buuren et. al. (1999), all variables in the analysis model were included as predictors (MOSAIC Consultant Ratings in 15-16 and 16-17, school, MOSAIC curriculum condition, race, and gender) with the exception of SDQ and DESSA, which were included only as dependent variables to reduce collinearity. As van Buuren et. al. (1999) point out, inclusion of all variables as predictors would theoretically give the most accurate imputed results. However, this is not feasible due to multicollinearity and computational problems. Thus, they recommend including 15-25 variables in multiple imputation. Thus, variables significantly correlated with SDQ and DESSA, or considered relevant on theoretical grounds were included as additional predictors (MOSAIC Student Curriculum Recall 2015-16 and 2016-17; Teacher Feedback Return Rate 2015-16 and 2016-17; grade level; accommodations, lunch status).

The five imputed data sets and complete cases were analyzed using the same model structures in R. This approach allowed for the inclusion of the full analysis sample of 308 students in both the intervention Year 1 and intervention Year 2 models, which in turn enabled a

more nuanced interpretation of program implementation quality impact over two years. Due to analytic limitations (e.g., imputation conduced on SPSS and MLM analysis conducted on R), pooled estimates were not able to be generated. Following guidelines from Sterne et. al. (2009), full predictor models were run on all six datasets, and results were compared. Results largely followed the same trends across datasets. Imputation 5 was selected as the most representative dataset (significant predictors in each model were found in more than half (≥ 3) of the six datasets), and thus model building and final model results presented and discussed in upcoming sections are from this imputation.

Results

Data Analysis Procedure

To assess research questions 1 and 2 multilevel models (MLM) were built using the lme4 package (https://github.com/lme4/lme4/) on R studio with student as level 1 and classroom as level 2. Separate models were built to assess four outcomes: (1) SEL in Spring Year 1, (2) SEL in Spring Year 2, (3) negative mental health in Spring Year 1 and (4) negative mental health in Spring Year 2. For each model, full maximum likelihood estimation for the fixed and random effects were applied (Singer & Willett, 2003, p. 88). For all models, level 2 predictors and covariates (MOSAIC Consultant Ratings in Year 1 and Year 2, school—dummy coded with B as reference—and MOSAIC Curriculum Condition) were entered as fixed slopes and level 1 predictors and covariates (Fall DESSA total scores, Fall SDQ mean scores, gender, and race) were entered as both random and fixed slopes and assessed for goodness of fit. Models were built from the ground up, with each new model with additional predictor compared to the previous model. If the difference between models was significant, the more complicated model (e.g., random and/or more predictors) was selected. If the difference between models was

nonsignificant, the simpler model (e.g., fixed and/or less predictors) was selected. This process was repeated for all predictors and covariates, and final models represent the best fit (Finch, Bolin, & Kelley, 2016). To assess Research Question 3, additional predictor (DESSA Year 1 and Year 2 Change Scores) and interaction (DESSA Change*MOSAIC Consultant Rating) were added to existing models to assess for mediation pathway of acquisition of SEL between program implementation quality and indicators of negative mental health in Year 1 and Year 2 (MacKinnon, Fairchild, & Fritz, 2007).

Models Predicting Teacher-rated SEL

Predicting teacher-rated SEL year 1 (Model 1). The null model, without predictors, was run to determine the intraclass correlation coefficient (ICC), or the proportion of variance attributable to differences among classrooms. The ICC of .30 indicates that 30% of variance in Spring Year 1 SEL is attributable to differences among classrooms. Next, Fall Year 1 SEL was added as a fixed and random predictor. The two models (fixed and random) were compared utilizing a chi-square test and no significant difference was found, so the variable was kept as fixed. The model with Fall Year 1 SEL as a fixed predictor accounted for more variance than the null (χ_2 (4) = 33.497 $p < .001$). This same process was repeated for all remaining predictors and covariates to determine the final model. Year 1 program implementation quality was not a significant predictor of Spring Year 1 SEL, and thus Hypothesis 1a was not supported (the variable was kept in the final model). Gender was a significant random predictor ($t = 4.92$, $p < .001$) such that female-identified students were more likely to have higher SEL at the end of year 1, when controlling for Fall Year 1 SEL. Race was also a significant fixed predictor ($t = 3.69$, $p < .001$), such that White and Asian students were more likely to have higher SEL than their counterparts of other racial identities (majority Hispanic and Black students). Students in

Schools A ($t = -3.22$, $p < .003$), and C ($t = -2.16$, $p < .04$) were more likely to have lower SEL

scores than School B. MOSAIC curriculum condition was not a significant predictor and did not

contribute to the model fit. It was, therefore, not included in the final model. The final model can

be found in Table 5:

$$SEL.SY1 = \beta_{00} + \beta_{01}*(SEL.FY1_i) + \beta_{02}*(PROG.IMP.Y1_i) + \beta_{03}*(GENDER_i) +$$

$$\beta_{04}*(RACE_i) + \beta_{05}*GENDER_i*CLASS.Y1_{ti} + \beta_{06}*(SCHOOL\ A_i) + \beta_{07}*(SCHOOL\ C_i) + r_{0i}$$

$$+ r_{1i}*CLASS.Y1_{ti} + e_{ti}).$$

Predicting teacher-rated SEL year 2 (Model 2). The null model, without predictors,

demonstrated an ICC of .42, indicating that 42% of variance in post-intervention year 1 social-

emotional skills is attributable to differences among classrooms. Fall Year 2 SEL was included

as a random predictor ($t = 3.30$ $p = .002$) in the final model. Gender ($t = 3.21$ $p = .001$) was a

significant fixed predictor, such that females were more likely to have higher SEL than males.

Race was also a significant fixed predictor ($t = 2.38$ $p = .018$), such that White and Asian

students were more likely to have higher SEL ratings than students from other racial/ethnic

backgrounds. Both predictors were included in the final model. Program implementation quality

in Year 2 was not a significant predictor of SEL Spring Year 2, and thus did not support

Hypothesis 1b (the variable was kept in the final model). School and MOSAIC curriculum

condition did not significantly predict SEL or contribute to the fit of the model and were,

therefore, not included in the model. The final model can be found in Table 6:

$$SEL.SY2 = \beta_{00} + \beta_{01}*(SEL.FY2_i) + \beta_{02}*(PROG.IMP.Y2_i)\ \beta_{03}*(GENDER_i) + \beta_{04}*(RACE_i)$$

$$+ \beta_{05}*\ SEL.FY2_{ii}*\ CLASS.Y2_{ti} + r_{0i} + r_{1i}*CLASS.Y2_{ti} + e_{ti}.$$

Models Predicting Self-Reported Negative Mental Health

Predicting self-reported negative mental health, year 1 (Model 3). The null model, without predictors, demonstrated an ICC of .03, indicating that 3% of variance in Spring Year 1 negative mental health is attributable to differences among classrooms. Although 3% is a relevantly low ICC, it is still large enough to support continued analysis utilizing multilevel modeling (Schoeman, 2019). Year 1 Fall negative mental health was added first as a random predictor and was significant ($t = 6.08$, $p < .001$) and a better fit than as a fixed predictor. Year 1 program implementation quality was added as the next fixed predictor and was not significant. Thus, Hypothesis 2a was not supported (variable was kept in final model). Gender, race, and MOSAIC curriculum condition were all found to be nonsignificant predictors that did not contribute to the goodness of fit for the model, and for this reason, they were not included in the final model. School was a significant fixed predictor, such that students in School A were more likely to have higher negative mental health in Spring Year 1 ($t = 3.46$, $p = .001$). The full model can be found in Table 7:

$$MEN.HEA.SY1 = \beta_{00} + \beta_{01}*(MEN.HEA.FY1_i) + \beta_{02}*(PROG.IMP.Y1_i)\ \beta_{03}*(SCHOOL_i) +$$
$$\beta_{04}*(SCHOOL_i) + \beta_{05}*(MEN.HEA.FY1_i)*CLASS.Y1_{ti} + r_{0i} + r_{1i}*CLASS.Y1_{ti} + e_{ti}.$$

Predicting self-reported negative mental health, year 2 (Model 4). The null model, without predictors, demonstrated an ICC of .11, indicating that 11% of the variance in Spring Year 2 negative mental health can be attributed to differences among classrooms. Fall Year 2 negative mental health was added as the first predictor and was found to be significant and contribute the best fit as a random predictor ($t = 5.18$, $p < .001$). Year 2 program implementation quality was added as a fixed predictor and was found to be significant ($t = -4.76$, $p < .001$) and to significantly contribute to the fit of the model (χ_2 (7) $= 82.81$ $p < .001$), indicating that students with more positive MOSAIC Consultant Ratings were more likely to report less negative mental

health. This finding supports Hypothesis 2b. Race was found to be a significant fixed predictor (t = -2.14, p = .034) and to contribute to the fit of the model (χ_2 (8) = 4.03, p = .04), indicating that White and Asian students were more likely to report less negative mental health compared to their peers of other racial backgrounds. School was also found to be a significant fixed predictor, such that students in School A were more likely to report less negative mental health in Spring Year 2 than students in School B (t = -2.91, p = .005). Gender and MOSAIC Curriculum condition did not significantly predict negative mental health or add to the fit of the model. They were, therefore, not included in the final model. The final model can be found in Table 8:

$$MEN.HEA.SY2 = \beta_{00} + \beta_{01}*(MEN.HEA.FY2_i) + \beta_{02}*(PROG.IMP.Y2_i) + \beta_{03}*(RACE_i) +$$

$$\beta_{04}*(SCHOOLA_i) + \beta_{05}*(SCHOOLC_i) + \beta_{06}*MEN.HEA.FY2_i*CLASS.Y2_{ti} + r_{0i} +$$

$$r_{1i}*CLASS.Y2_{ti} + e_{ti}).$$

Models Evaluating Change in Teacher-rated SEL as Mediator

  Change in teacher-rated SEL as mediator of self-reported negative mental health, year 1 (Model 5). The change in SEL was computed by utilizing the unstandardized residual change scores of Spring Year 1 DESSA regressed on Fall Year 1 DESSA**.** Following procedures indicated by MacKinnon, Fairchild, & Fritz (2007), the mediation analysis was done in several steps. First, Year 1 program implementation quality was tested as a predictor of both Year 1 change in SEL and Spring Year 1 negative mental health, controlling for Fall negative mental health. Both analyses yielded nonsignificant results. Next, Year 1 change in SEL was evaluated as a predictor of Spring Year 1 negative mental health and was found to be significant (β_{02} = -.008, p < 0.001). Finally, all predictors and the interaction of change in social-emotional skills were tested in one model, along with covariates, and the interaction term (Year 1 program implementation quality*change in SEL) was not found to be significant. Thus, Year 1 change in

SEL did not mediate the relationship between Year 1 program implementation quality and Year 1 negative mental health. Hypothesis 3a was not supported. The final model can be found in Table 9.

Change in teacher-rated SEL as mediator of self-reported negative mental health, year 2 (Model 6). The change in SEL was computed by utilizing the unstandardized residual change scores of Spring Year 2 DESSA regressed on Fall Year 2 DESSA. Again following the procedures indicated by MacKinnon, Fairchild, and Fritz (2007), the mediation analysis was done in the same steps. First, Year 2 program implementation quality was tested as a predictor of Year 2 change in SEL and was found to be nonsignificant ($\beta_{01} = -1.53$). Year 2 program implementation quality was a significant predictor of Spring Year 2 negative mental health, controlling for Fall negative mental health ($\beta_{03} = -.11**$). Year 2 change in SEL was then tested as a predictor of Spring Year 2 negative mental health and was found to be nonsignificant. Finally, all predictors and the interaction term (Year 2 program implementation quality*change in Year 2 SEL) were tested in one model, along with covariates, and the interaction term was not found to be significant. Thus, Year 2 change in SEL did not mediate the relationship between Year 2 program implementation quality and Year 2 mental health. Hypothesis 3b was not supported. The final model can be found in Table 10.

Post-hoc Analysis

To better understand the relationship between teacher-rated SEL, self-reported negative mental health, and program implementation quality, two exploratory multilevel models were built utilizing the same procedure as earlier models in this analysis (using R studio and lme4 package). The first model (Model 7) evaluated Fall Year 1 SEL as a moderator of Year 1 program implementation quality in predicting negative mental health Spring Year 1. Gender and

School were kept in the model as significant covariates. Although the interaction of SEL and program implementation quality was not a significant predictor of Spring Year 1 negative mental health at a .05 level, the variable approached significance ($t = -1.60$, $p = .11$). See Figure 5 for the interaction plot. The full model is depicted in Table 9:

$$MEN.HEA.SY1 = \beta_{00} + \beta_{01}*(SEL.FY1_i) + \beta_{02}*(PROG.IMP.Y1_i) + \beta_{05}*(GENDER_i) +$$
$$\beta_{04}*(SCHOOLA_i) + \beta_{05}*(SCHOOLC_i) + \beta_{06}*(PROG.IMP.Y1_i)*(SEL.FY1_i) + r_{0i} +$$
$$r_{1i}*CLASS.Y1_{ti} + e_{ti}.$$

Model 8 evaluated Fall Year 2 SEL as a moderator of Year 2 program implementation quality in predicting negative mental health in Spring Year 2. Race ($t = -2.98$, $p = .003$) and school were kept in the model as significant predictors. As an individual predictor, Fall Year 2 SEL was significantly related to Spring Year 2 negative mental health scores in the unexpected direction, such that higher SEL predicted higher negative mental health ($t = 2.25$, $p = .03$). However, when included as an interaction term with program implementation quality, the relationship flipped to the expected direction. Thus, when the interaction of SEL and program implementation quality was higher, students were significantly more likely to have less negative mental health in the Spring of Year 2 ($t = -2.31$, $p = .02$). See Figure 6 for the interaction plot. The full model 8 is depicted in Table 10:

$$MEN.HEA.SY2 = \beta_{00} + \beta_{01}*(SEL.FY2_i) + \beta_{02}*(PROG.IMP.Y2_i) + \beta_{03}*(RACE_i) +$$
$$\beta_{04}*(SCHOOLA_i) + \beta_{05}*(SCHOOLC_i) + \beta_{06}*(PROG.IMP.Y2_i)*(SEL.FY1_i) + r_{0i} +$$
$$r_{1i}*CLASS.Y2_{ti} + e_{ti}.$$

Discussion

Findings from this analysis help to illuminate the relationship of program implementation quality of the MOSAIC SECD intervention, (teacher-rated) student social-emotional learning,

and (self-reported) student negative mental health. Importantly, program implementation quality significantly predicted student negative mental health in Spring Year 2, when controlling for Fall Year 1. Although change in social-emotional learning did not significantly mediate the relationship between program implementation quality and negative mental health for Year 1 or Year 2 as was predicted, additional post-hoc analysis revealed that Fall SEL moderated program implementation quality as a predictor of Spring indicators of negative mental health for both years (approaching significance for Year 1, $p = .11$ and significantly for Year 2, $p = .02$). With regard to covariates, gender and race were significant predictors of SEL increases in both Year 1 and Year 2, such that female and White or Asian-identified students (respectively) were likely to have higher SEL in the Spring when controlling for Fall scores. School context was also a predictive factor. Students in School A had significantly lower growth in SEL and higher increase in negative mental health than students in School B in Year 1. Additionally, in both Year 1 and Year 2, a large percent of variance of student SEL was accounted for by classroom level characteristics (Year 1 ICC = .30 and Year 2 ICC = .42). By contrast, a smaller percentage of variance in student negative mental health was accounted for by classroom level characteristics (Year 1 ICC = .03 and Year 2 ICC = .11). These findings, their implications, study limitations, and future directions are explored in this section.

Program Implementation Predicting Mental Health Year 2

The literature from the field reflects that implementation of SECD interventions in schools impacts student outcomes, including student mental health and emotional distress (e.g., Askell-Williams et al., 2013; Durlak et al., 2011; Durlak & Dupree, 2008; Ransford et al., 2009). However, we are still in the early stages of understanding the differential impacts of the dimensions of implementation (Low, Smolkowski, Cook, 2016). There are indications that

program implementation quality, as measured by an individual familiar with both intervention and context (e.g., coach or consultant), can predict student outcomes (Reyes et al., 2012). Thus, it is well supported that program implementation quality would predict more positive student indicators of mental health in the current study.

It is possible that this finding developed in Year 2 (and not Year 1) due to increased teacher comfort with the intervention. In a similar study evaluating the implementation of an SECD program in nine urban schools in Canada, made up of multi-ethnic, predominantly from socio-economically disadvantaged families, the majority of teachers reported that over the course of the intervention, they improved in their pedagogical methods of delivering the program (Ayotte, Saucier, Bowen, Laurendeau, Fournier, & Blais, 2003). It is possible that by year two, teachers had gains in their pedagogical capacity that contributed to significant student outcomes in the second but not first year. In fact, in the Durlak et al. (2011) meta-analysis, the strongest findings and those usually cited were those obtained after two years of systematic SEL. The interpretation typically is made that the students are not truly grasping the skills until the second year, but another interpretation would be that it is not until the second year that most teachers begin to become comfortable with and knowledgeable about SEL interventions, leading to improved instruction and better student outcomes (Durlak et al., 2011).

Many have pointed out that school and community interventions often take more than one year to produce results. Renowned methodologist Don Campbell is quoted as saying, "You should evaluate no program before it is proud." Legendary community psychologist, Seymour Sarason, was among the first to codify the "three-year rule" for school interventions, advising researchers to expect minimal outcomes until three years of intervention (Romasz, Kanto, & Elias, 2004, p. 101). Perhaps due to overall program development on the part of the school,

teacher, and student, the relationship between program quality implementation and student mental health became pronounced in Year 2.

Fall SEL as a Moderator

We can further understand the relationship of program implementation quality and student negative mental health outcomes by considering students' initial SEL. This study found that Fall SEL moderated program implementation quality as a predictor of Spring negative mental health for both study years (significance approached significance at $p = .11$ for Year 1 and $p = .02$ for Year 2). This finding suggests that social-emotional competencies do—as other literature suggests (e.g., Guerra & Bradshaw, 2008; Lewis et al., 2013; Modecki, Zimmer-Gembeck, & Guerra, 2017; and Taylor et al., 2017)—relate to mental health-related outcomes of SECD interventions. However, in this case, it seems that only when students entered the MOSAIC classroom with a certain level of SEL were they able to experience gains in their mental health through engaging with the intervention. To this researcher's knowledge, no other published studies have evaluated this.

Many studies in complex settings in which communities and individuals face challenges such as poverty, racial/ethnic discrimination, increased exposure to trauma, and other impacts of systemic marginalization have found that prevention efforts often show success through the slowing down of negative trajectories (e.g., Weissberg, Jackson, & Shriver, 1993). Considering this background, it may be the case that in low-resourced urban schools, beset by many of the challenges outlined above, students have not always been given the support and opportunity to cultivate their SEL prior to middle school. And, given the added challenges facing students in these circumstances, students with lower SEL in the Fall may need additional supports and interventions in order to benefit from the universal SECD intervention. Alternatively, this finding

might support the need for earlier SEL intervention being important in order to be ready and able to engage with the MOSAIC intervention.

Gender and SEL

Prior research indicates that female participants may be more likely to have higher interpersonal social-emotional competencies than their male counterparts. Rowe and Trickett (2017) found that 37 of the studies in the Durlak et al. (2011) meta-analysis evaluated gender as a moderator for program outcomes. Although many studies did not find gender to be a significant moderator ($n = 22$), one trend identified female participants as having higher interpersonal effectiveness outcomes than their male counterparts. Considering that three out of the eight items assessing SEL for the current study explicitly connect to interpersonal skills ("Accept responsibility for what he/she did", "do something nice for somebody", and "contribute to group efforts"), it is possible that females demonstrated higher SEL at the end of each intervention year due to stronger interpersonal skills. If we are to understand that baseline SEL ability is an important factor in MOSAIC intervention outcomes, it may be that males are more likely to be lower in SEL, and could particularly benefit from additional supports prior and during intervention.

Race, SEL, and Negative Mental Health

Prior research on racial difference in SEL has produced diverse findings. Garner et al. (2014), pointed to one finding that African American students were rated lower on SEL than their non-African American counterparts (Elias & Haynes, 2008) and another that low-income urban African American children were more sympathetic to severe peer difficulties (e.g., parental abuse) and less sympathetic to less severe peer difficulties (Leff et. al., 2012). Garner and colleagues went on to explain that for Hispanic students, as noted previously, a population

very rarely studied in the context of SEL, there has been some indication of higher risk for internalizing problems (Varela et al., 2009), and also differential understandings of role of acculturation and social-emotional competencies (2014).

Of particular relevance in the current study, and likely in many of the studies cited above, is that student SEL was rated by teachers. There is substantial literature in the education field indicating that students of color, and Black students in particular, are subject to disproportionately high rates of disciplinary referrals (e.g., Gregory, Skiba, & Noguera, 2010; Skiba et al., 2002). Research indicates that negative beliefs and expectations held by teachers (Ferguson, 2000) and "societal stereotypes, implicit bias, or cultural mismatch" between teachers and students (Gregory, Skiba, & Noguera, 2010) contribute to the disproportionally high disciplinary referrals for students of color. It is quite possible that the same forces influencing school discipline referrals also influence teacher ratings of students' social-emotional competencies and behaviors. In her assessment of the field of SEL, Hoffman (2009) warned against the potential for SEL becoming "another way to focus attention on measurement and remediation of individual deficits rather than a way to redirect educators' focus toward the relational contexts of classrooms and schools" (p. 533). Although Hoffman's words echo across the many analyses that have focused on student individual outcomes related to SECD interventions, they seem particularly pertinent to findings related to race. Such findings must be interpreted within a relational context, considering the strong forces of systemic racism, prejudice, and bias that permeate the urban settings, schools, and the teacher-student relationship in the United States. Thus it is through this lens that findings indicating White and Asian students exhibited higher SEL than the Black, Hispanic and other racial/ethnically identified counterparts.

With regard to negative mental health, measured by student self-report, race was not a significant covariate in Year 1. However, in Year 2, Asian and White students were found to have lower scores on indicators of negative mental health in the Spring than students of other racial backgrounds. Although recent epidemiological studies of adolescent mental health disorders in the United States did not reveal differential prevalence rates by race or ethnicity (Kessler et al., 2011), this finding may be considered through 'The Conceptual Model for Child Mental Health and Mental Health Service Disparities' (Alegría et al., 2015), discussed in the Cultural and Ecological Considerations section of this paper. The Model posits that the interconnected network of risk factors relating bi-directionally to youth and disproportionately affecting youth of color contributes to the development of mental health disorder and the higher rates of persistent mental health disorder for individuals of color in adulthood.

The Importance of School Context

For the majority of completed analyses related to SEL and negative mental health student outcomes, school was a significant covariate. This finding underscores the larger point that interventions are context-dependent, and MOSAIC is no exception. Factors at the school level—including school climate; teacher turnover; administrator style and support of the MOSAIC intervention; and MOSAIC implementation structure—no doubt impacted student outcomes. From the early definition and inception of SEL school-wide interventions, it has been understood that, "a context that is caring, supportive and challenging also leads to better SEL outcomes" (Elias et al., 1997, p. 75).

Gager and Elias (1997) conducted a survey of successful implementation components for prevention programs in high and low-risk settings across all public school districts in the State of New Jersey ($n = 311$ respondents). In this study, high-risk settings were characterized as those

with lowest scores on the District Factor Group, a socioeconomic/ demographic index developed by the New Jersey State Department of Education. Successful, or "resilient," programs in high-risk settings had strong Program Institutionalization and Implementer and Recipient Focused Practices. Key factors of Program Institutionalization include: a strong program support team or individual; a sense of ownership over the program; high visibility of the program; formal and informal training procedures; and consistent structure and adherence to the program. Implementer and Recipient Focused Factors include: engaging and varied materials; linkage to existing school/district goals; and consistent administrator support. Authors found that the same evidence-based programs were found to be successful or unsuccessful in implementation, depending on context (Gager & Elias, 1997).

Gathering detailed data from a large enough *n* of schools in order to analyze school-level factors in relation to student-level program outcomes poses an enormous obstacle. As an alternative, exploring school-level factors through a more qualitative, case-study lens can provide essential information about how school context may have influenced student outcomes. In another urban school district in New Jersey serving predominantly students of color from low-income households, implementers of a district-wide SECD program explained the importance of factors such as administrative support of the intervention at the district and school level, teacher familiarity and buy-in of the program, and intervention follow-through so that SECD becomes integral to the culture and climate of the school (Romasz, Kantor, & Elias, 2004).

In the three schools involved in the current study, school-level implementation factors related to MOSAIC differed over the course of the two years of intervention. A full recount of each school's MOSAIC implementation journey requires its own paper. However, several key points may help illustrate differences across school context that may contribute to the current

findings. Due to its size, the largest school in the study by far, MOSAIC Consultants in School A, the largest school in the study by far, struggled to meet, observe, and support all teachers. Further, this school was labeled a "Focus" school by the State Department of Education, indicating that student growth and achievement scores were historically subpar. This label came with added curricular and instructional demands, oversight, and resources, all framed in the context of extra pressure to improve academic performance. For instance, School A was implementing an intensive math curriculum planned down to the minute, and thus MOSAIC classes were scheduled for the first 15 minutes of every first period with the exception of math classes. Students with math first period had MOSAIC for the first 15 minutes of the Language Arts class later in the day. However, there was a frequent presence of literacy coaches in the school and when they showed up in Language Arts classes, planned MOSAIC lessons would be set aside. The many competing demands and pressures, coupled with the size of school were among many complex school factors at School A.

In School C, a smaller K-8 school, administrator support for MOSAIC was ambivalent throughout the two years of intervention. This lead to fluctuating MOSAIC intervention structure, initially scheduling MOSAIC for morning homeroom and later changing MOSAIC to take place as part of the social studies curriculum. Additionally, several MOSAIC teachers were transferred to different grades and subject areas over the course of the study. School B, also a smaller K-8 school, had strong administrator support for MOSAIC, and the principal demonstrated interest in conducting his own analyses of MOSAIC impact. The MOSAIC Consultant, who worked with the school for both study years, was consistently welcomed to observe MOSAIC lessons, and a strong MOSAIC support team was established within the school. MOSAIC was consistently scheduled during the morning homeroom class period across

the entire intervention. This summary by no means captures the complicated nuances of the MOSAIC intervention. Rather, the information shared demonstrates that broadly speaking, School B conducted stronger school-level implementation of MOSAIC.

Suggestions for Future Research and Practice

Results and limitations of the current study provide insight into important suggestions for future research and practice. Suggestions for future research include: making improvements to the consultant rating measure; utilizing methods to encourage greater consultant access to intervention classrooms; measuring additional dimensions of implementation; evaluating program implementation outcomes at later stage in program development; and evaluating additional student outcome variables. Future practitioners of MOSAIC may consider utilizing the first intervention year to lay a strong ecological foundation for school-wide implementation and providing extra support and skill development opportunities for students with lower SEL at start of MOSAIC intervention. These suggestions will be elaborated below.

The MOSAIC Consultant Rating measure had several limitations that could be addressed to increase measure validity in future studies. First, it is likely that MOSAIC Consultant Ratings abided by some level of internal norming based on the contextual and individual factors the consultant experienced in their intervention school(s). In addition to the consultants' group note sharing and supervision, in which group norms were established for program implementation quality, additional steps could help create an "absolute" program implementation quality rating across schools. For example, future interventions could encourage consultants to spend time whenever possible in other intervention schools to help provide a common calibration of ratings. Another helpful action could be for consultants to review videos of the standard intervention to create benchmark rating standards (training videos for the MOSAIC intervention are now

available on the website, https://www.secdlab.org/training-video, and could be used for this purpose). Having multiple consultants at each intervention school conducting ratings would also strengthen the measure and provide opportunity for inter-rater reliability analysis.

In the current study, a substantial number of classrooms were excluded because MOSAIC consultants did not have enough information to rate the program implementation quality. Repeatedly over the course of the study, consultants reported (in notes and supervision meetings) barriers to observing classrooms, namely teacher discomfort with observation coupled with lack of explicit support from administrators for consultant classroom visits. As noted earlier, in School A, observations around academic subject area teaching were frequent and seen as highly evaluative, an attitude that extended to any teacher observation, including MOSAIC. As a result, important information about program implementation quality was not captured by the current study. One suggestion for future researchers is to establish explicit agreements for consultant/coach observations and interactions with teachers. This, of course, is easier said than done, especially in the context of distressed urban middle schools, where teacher observation, evaluation, and student achievement has become increasingly used to inform teacher job security, salary, and remediation (Noguera & Fine, 2011). In such an environment, a teacher's concern or hesitation to allow outside observers into one's classroom is understandable. However, some methods that could help increase teacher openness include: rapport and relationship building between consultant and teachers; smaller consultant-teacher ratios (e.g., no greater than 1:20) to help enable relationship building; explicit support from administration for consultant visits and observations; integration of teacher feedback into ongoing program development.

Another approach to capture more program implementation information is to gather data on multiple dimensions of implementation. From the nascent pool of empirical evaluations of the impact of different implementation dimensions on SECD program outcomes, there is evidence suggesting that multiple dimensions, and their interaction, may be important predictors. For instance, Reyes et al. (2012) found that program quality was only a significant predictor of program outcomes when interacting with dosage and training dimensions. Additionally, in a study of 61 schools, 321 teachers, and over 7,300 early elementary school students in the Puget Sound area of Washington—ranging from urban to rural—researchers assessed the following dimensions of implementation of the Second Step SEL program in relation to student outcomes: adherence, dosage, student engagement, and generalization. Researchers found that student engagement was the most critical dimension in predicting student outcomes (Low, Smolwolski, & Cook, 2016). By evaluating multiple dimensions of implementation, in particular student engagement and implementation quality, and their interaction variables, in other studies in similar under-resourced urban schools, researchers can build a more nuanced understanding of the role of implementation in SECD intervention student outcomes of SECD interventions in these settings. Additionally, collecting multiple dimensions could open up opportunity for including more participants and potentially using multiple imputation methods (e.g., utilizing other teacher and classroom variables to impute for missing ratings) to capture a larger sample and reduce potential for bias in results.

Another consideration for future research is the timing for program outcome measurement. It may be beneficial to allow the intervention one year to develop at the school before expecting any changes in student variables. Results of the current study indicated that student SEL and negative mental health was not predicted by program implementation quality in

Year 1 of the MOSAIC intervention. A plethora of school and community research has indicated that results in the first year of intervention are unlikely (e.g., Don Campbell). It may be the case that in under-resourced urban middle schools, facing many and varied challenges, the timeline for intervention results may be expanded beyond the second year or even Seymour Sarason's "three-year rule" to a longer timeframe of 5-7 years (Linsky & Elias, 2018). Although this point has been made repeatedly in the literature, it is essential that grant-funders allow for this initial development year within proposed project structures.

Along with considering an elongated timeframe for program results, researchers in urban, low-resource middle schools may also consider additional outcome measures beyond student competencies and performance (Hoffman, 2009). In addition to measuring student SEL and negative mental health, measures of student perception of school climate, student purpose and character virtues—particularly those understood as related to cultivating resilience in the face of trauma and adversity—and student social support may also capture integral processes in effective SECD interventions in these settings.

Practitioners of SECD interventions (and MOSAIC in particular) in similar contexts may consider assessing student SEL before starting the intervention. For students demonstrating lower SEL, it may be important to offer additional supports (e.g., small skill groups) either prior or concurrent to the SECD intervention. In facing the challenges discussed in the current study, it may also be especially important to provide SECD interventions in elementary school to enable students to partake in interventions, such as MOSAIC in middle school, when serving populations of students who have experienced disproportional amounts of community and personal trauma and adversity (e.g., Osher et al., 2016). In sum, given the many complexities, challenges, and obstacles to implementation of SECD interventions in the urban middle school, it

follows that additional supports, preparation, and dosage might be necessary for students to benefit fully and feel improvements in their SEL and mental health

Conclusion

The momentum for schools to implement school-wide SECD interventions is moving forward with full speed (e.g., Dusenbury et al., 2015; Jones & Kahn, 2017). Tempering this general conclusion will be essential to SECD fulfilling its primary potential. Key points related to that tempering, reviewed above, follow. Several major meta-analyses have provided empirical support that SECD interventions promote positive student outcomes, including increase in SEL and decrease in negative mental health (Durlak et al., 2011; Sklad et al., 2012; Taylor et al., 2017). Some evidence suggests that implementation of SECD programs predicts student outcomes (e.g., Durlak, 2016; Durlak et al., 2011). In particular, program implementation quality seems to be an important dimension of implementation in predicting student outcomes (Reyes et al., 2012). Theory also suggests that in these high-risk contexts, it is essential that SECD interventions help students cultivate both the virtues necessary for resilience in the face of adversity and trauma and the SEL to enact these virtues. However, although we know that interventions are different across contexts, and that high-risk contexts present unique challenges to implementation (Gager & Elias, 1997), there has not been much research evaluating implementation dimensions of SECD in low-resource, urban middle school setting as predictors of student outcomes.

The current study investigated whether program implementation quality of the MOSAIC intervention in three urban middle schools predicted student SEL and negative mental health over the course of Year 1 and Year 2 of the intervention. Results demonstrated that the better program implementation quality of MOSAIC predicted decreased negative mental health in Year

2 of the intervention. Results also indicated that students with higher SEL at the start of each school year were more able to benefit from MOSAIC, and had less indicators of negative mental health than their counterparts who started the school year with lower SEL. These findings support the continued implementation of SECD interventions in low-resource urban middle schools; underscore the challenges in implementing and evaluating programs in these settings and potential need for increased dosage and supports for students to fully benefit from programs; and underscore the need for future research expanding our understanding of program implementation, student, and school/community outcomes.

References

AEI/Brookings Working Group on Poverty and Opportunity. (2015). Opportunity, responsibility, and security: *A consensus plan for reducing poverty and restoring the American dream.* AEI.

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

Askell-Williams, H., Dix, K. L., Lawson, M. J., & Slee, P. T. (2013). Quality of implementation of a school mental health initiative and changes over time in students' social and emotional competencies. *School Effectiveness and School Improvement*, *24*(3), 357-381.

Ayotte, V., Saucier, J. F., Bowen, F., Laurendeau, M. C., Fournier, M., & Blais, J. G. (2003). Teaching multiethnic urban adolescents how to enhance their competencies: Effects of a middle school primary prevention program on adaptation. *Journal of Primary Prevention*, *24*(1), 7-23.

Buchanan, R., Gueldner, B. A., Tran, O. K., & Merrell, K. W. (2009). Social and emotional learning in classrooms: A survey of teachers' knowledge, perceptions, and practices. *Journal of Applied School Psychology*, *25*(2), 187-203.

Campbell, D. T. (1988). Evolutionary epistemology. In E. S. Overman (Ed.), *Methodology and epistemology for social science: Selected papers of Donald T. Campbell* (pp. 393-434). Chicago: University of Chicago Press.

Collaborative for Academic, Social, and Emotional Learning [CASEL]. (2012). *2013 CASEL guide: Effective social and emotional learning programs, preschool and elementary school edition*. KSA-Plus Communications, Inc.

Castro-Olivo (2014). Promoting social-emotional learning in adolescent Latino ELLs: A study of the culturally adapted strong teens program. *School Psychology Quarterly, 29(*4), 567-577. .

Dane, A. V. & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18*, 23-45.

Domitrovich, C.E., Syvertsen, A. K. & Calin, S. S. (2017). *Promoting social and emotional learning in the middle and high school years.* Edna Bennett Pierce Prevention Research Center, Pennsylvania State University.

Durlak, J. A. (2016). Programme implementation in social and emotional learning: basic issues and research findings. *Cambridge Journal of Education*, *46*(3), 333-345.

Durlak, J.A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41,* 327-50.

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.

Dusenbury, L.A., Zadrazil Newman, J., Weissberg, R.P., Goren, P., Domitrovich, C.E., & Mart, A.K. (2015). The case for preschool through high school state learning standards for SEL. In J.A. Durlak, C.E. Domitrovich,R.P. Weissberg, & T.P. Gullotta (Eds.), *Handbook for social and emotional learning: Research and practice* (pp. 532-548). New York, NY: Guilford Press.

Elias, M. J. (2014). The future of character education and social-emotional learning: The need for whole school and community-linked approaches. *Journal of Research in Character Education*, *10*(1), 37-42.

Elias, M. J., Bruene-Butler, L., Blum, L., & Schuyler, T. (2000). Voices from the field: Identifying and overcoming roadblocks to carrying out programs in social and emotional learning/emotional intelligence. *Journal of Educational and Psychological Consultation*, *11*(2), 253-272.

Elias, M. J., & Haynes, N. M. (2008). Social competence, social support, and academic achievement in minority, low-income, urban elementary school children. *School psychology quarterly*, *23*(4), 474.

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

Ferguson, A. A. (2000). *Bad boys: Public school and the making of Black masculinity*. Ann Arbor: University of Michigan Press.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics.* California, U.S.: SAGE Publications Inc.

Finch, W. H., Bolin, J. E., & Kelley, K. (2016). *Multilevel modeling using R*. CRC Press.

Fixsen, D. L., Blase, K. A., Naoom, S. F., & Wallace, F. (2009). Core implementation components. *Research on Social Work Practice*, 19(5), 531-540.

Frey, K. S., Hirschstein, M. K., Snell, J. L., Van Schoiack Edstrom, L., MacKenzie, E. P., & Broderick, C. J. (2005). Reducing playground bullying and supporting beliefs: an experimental trial of the Steps to Respect program. *Developmental Psychology*, 41, 479–491.

Gager, P. J., & Elias, M. J. (1997). Implementing prevention programs in high-risk environments: Application of the resiliency paradigm. *American Journal of Orthopsychiatry*, 67(3), 363-373.

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

Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 38, 581-586.

Gorski, P. C., & Covert, B. (2000). A working definition of multicultural education. *Multicultural Pavilion*, 1-4.

Gottfredson, D. C., & Gottfredson, G. D. (2002). Quality of school-based prevention programs: Results from a national survey. *Journal of research in crime and delinquency*, *39*(1), 3-35.

Guerra, N. G., & Bradshaw, C. P. (2008). Core competencies to prevent problem behaviors and promote positive youth development. *New Directions in Child and Adolescent Development.* San Francisco, CA: Josey-Bass.

Greenberg, M. T., Weissberg, R. P., O'Brien, M. U., Zins, J. E., Fredericks, L., Resnik, H., & Elias, M. J. (2003). Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. *American Psychologist, 58*(6-7), 466-474.

Gregory, A., Skiba, R. J., & Noguera, P. A. (2010). The achievement gap and the discipline gap: Two sides of the same coin? *Educational Researcher*, *39*(1), 59-68.

Harlacher, J. E., & Merrell, K. W. (2010) Social and emotional learning as a universal level of student support: Evaluating the follow-up effect of strong kids on social and emotional outcomes. *Journal of Applied School Psychology*, *26*(3), 212-229.

Hatchimonji, D., Linsky, A., DeMarchena, S., Kim, S., Nayman, S., & Elias, M. (2017). Building a culture of engagement through participatory feedback processes. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 91(2), 59-65*.

Hatchimonji, D.R., Linsky, A.V., Elias, M. J. (2017). Cultivating noble purpose in urban middle schools: A missing piece in school transformation. *EDUCATION*. *138*(2) 162-178.

Jones, S. & Kahn, J. (2017). The evidence base for how we learn: Supporting students' social, emotional, and academic development. *The Aspen Institute, National Commission on Social, Emotional, and Academic Development.*

Kessler, R. C., Avenevoli, S., Costello, E. J., Georgiades, K., Green, J. G., Gruber, M., … Merikangas, K. R. (2012). Prevalence, persistence, and sociodemographic correlates of DSM-IV disorders, in the national comorbidity survey replication adolescent supplement. *Arch Gen Psychiatry, 69*(4), 372-380. DOI:10.1001/archgenpsychiatry.2011.160.

Lee, F. S., Heimer, H., Giedd, J. N., Lein, E. S., Šestan, N., Weinberger, D. R., & Casey, B. J. (2014). Adolescent mental health—opportunity and obligation. *Science*, *346*(6209), 547-549.

Leff, S. S., Lefler, E. K., Khera, G. S., Paskewich, B., & Jawad, A. F. (2012). Preliminary examination of a cartoon-based hostile attribution bias measure for urban African American boys. *American Journal of Community Psychology*, 49, 332–346.

Lewis, K. M., DuBois, D. L., Bavarian, N., Acock, A., Silverthorn, N., Day, J., Ji, P., Vuchinich, S., Flay, B. (2013). Effects of positive action on emotional health of urban youth: A cluster-randomized trial. *Journal of Adolescent Health, 53,* 706-711.

Linsky, A. V. & Elias, M. J. (2018). *Using social-emotional learning and character development for school turnaround: Principles exemplified in case examples*. Manuscript submitted for publication.

Linsky, A.V., Hatchimonji, D.R., Kruzik, C.L., Kifer, S., Franza, N., McClain, K., Nayman, S.J., & Elias, M.J. (in press). Students taking action together (STAT): Social action in urban middle schools. *Middle School Journal.* DOI: :10.1080/00940771.2018.1488469

Linsky, A., Hatchimonji, D., Vaid, E., Nayman, S., & Elias, M. (October 2018). The Importance of School Context in Understanding Student Character Development. Poster presented at SRCD 2018 Special Topic Meeting: Promoting Character Development Among Diverse Children and Adolescents: The Roles of Families, Schools, and Out-Of-School-Time Youth Development Programs in Philadelphia, PA.

Low, S., Smolkowski, K., & Cook, C. (2016). What constitutes high-quality implementation of SEL programs? A latent class analysis of Second Step® implementation. *Prevention Science*, *17*(8), 981-991.

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review Psychoogy*, *58*, 593-614.

Mahoney, J. L., Durlak, J. A., & Weissberg, R. P. (2018). An update on social and emotional learning outcome research. *Phi Delta Kappan*, *100*(4), 18-23.

Merikangas, K. R., He, J., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., … Swendsen, J. (2010). Lifetime prevalence of mental disorders in US adolescents: Results from the national comorbidity study-adolescent supplement (NCS-A). *Journal of the American Academy of Child and Adolescent Psychiatry*, *49*(10), 980–989. DOI: 10.1016/j.jaac.2010.05.017

Merrell, K. W. (2010). Linking prevention science and social and emotional learning: The Oregon resiliency project. *Psychology in The Schools*, *47*(1), 55-70.

Modecki, K. L., Zimmer-Gembeck, M. J., & Guerra, N. (2017). Emotion regulation, coping, and decision-making: Three linked skills for preventing externalizing problems in adolescence. *Child Development*, 88, 417-426.

Noguera, P. & Fine, M. (2011, May 9). Teachers aren't the enemy. *The Nation*. Retrieved from: www.thenation.com/article/teachers-arent-enemy/.

O'Connell, M. E., Boat, T., & Warner, K. E. (2009). *Preventing mental, emotional, and behavioral disorders among young people: Progress and possibilities* (Vol. 7). Washington, DC: National Academies Press.

Osher, D., Kidron, Y., Brackett, M., Dymnicki, A., Jones, S., & Weissberg, R. (2016). Advancing the science and practice of social and emotional learning: Looking back and moving forward. *Review of Research in Education, 40*, 644-681.

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.

Romasz, T. E., Kantor, J. H., & Elias, M. J. (2004). Implementation and evaluation of urban school-wide social–emotional learning programs. *Evaluation and Program Planning, 27*(1), 89-103.

Rowe, H. L., & Trickett, E. J. (2017). Student Diversity Representation and Reporting in Universal School-Based Social and Emotional Learning Programs: Implications for Generalizability. *Educational Psychology Review*, 1-25.

Rubin, D. B. (1987). Multiple Imputation for Nonresponse in Surveys. New York: Wiley.

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

Schoeman, A. (2019). Materials to accompany Schoemann (stats camp 2019) workshop on multilevel modeling in R [PowerPoint and sample R code]. Retrieved from http://marlab.org/Workshops/.

Simmons, D. N., Brackett, M. A., & Adler, N. (2018). Applying an Equity Lens to Social, Emotional, and Academic Development. The Pennsylvania State University.

Singer, J. D., Willett, J. B., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, Y: Oxford University Press.

Sklad, M., Diekstra, R., Ritter, M. D., Ben, J., & Gravesteijn, C. (2012). Effectiveness of school-based universal social, emotional, and behavioral programs: Do they enhance students' development in the area of skill, behavior, and adjustment? *Psychology in the Schools, 49*(9), 892-909.

Skiba, R. J., Michael, R. S., Nardo, A. C., & Peterson, R. L. (2002). The color of discipline: Sources of racial and gender disproportionality in school punishment. *Urban Review*, *34*, 317–342.

Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., ... & Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *The BMJ*, *338*, b2393. https://doi.org/10.1136/bmj.b2393.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics* (6th Edition). Boston, Massachusetts: Pearson.

Taylor, R. D., Oberle, E., Durlak, J. A., & Weissberg, R. P. (2017). Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. *Child Development, 88*(4), 1156-1171.

Urban, J. B., Hargraves, M., & Trochim, W. M. (2014). Evolutionary evaluation: Implications for evaluators, researchers, practitioners, funders and the evidence-based program mandate. *Evaluation and Program Planning, 45*, 127-139.

van Buuren, S., Boshuizen, H. C., & Knook, D. L. (1999). Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine*, *18*(6), 681-694.

Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy pro- tocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology, 61,* 620– 630. DOI:10.1037/0022–006x.61.4.620

Weissberg, R. P., Durlak, J. A., Domitrovich, C. E., & Gullotta, T. P. (2015). Social and emotional learning: Past, present, and future. In J. A. Durlak, C. E. Domitrovich, R. P. Weissberg, & T. P. Gullotta (Eds.), *Handbook of social and emotional learning: Research and practice* (pp. 3-19). New York, NY: The Guilford Press.

Weissberg, R. P., Jackson, A. S., & Shriver, T. P. (1993). Promoting positive social development and health practices in young urban adolescents. Social decision making and life skills development: *Guidelines for Middle School Educators*, 45-77.

Table 1

Demographic Characteristics of Original Sample by School, Year 1

	A	B	C	D	Total	χ^2
Female-Identified *n* (%)d	254 (48.8)	91 (48.7)	110 (44.7)	70 (48.6)	525	1.22
Male-Identified *n* (%)	267 (51.2)	96 (51.3)	136 (55.3)	74 (52.4)	573	-
White *n* (%)	51 (9.7)	51 (27.2)	18 (7.3)	8 (5.6)	128	55.75***
Black *n* (%)	163 (31.3)	33 (17.6)	19 (7.7)	78 (54.2)	293	114.24***
Hispanic *n* (%)	201 (38.6)	82 (43.9)	164 (66.7)	41 (28.5)	488	71.36***
Asian *n* (%)	103 (19.8)	19 (10.2)	45 (18.3)	12 (8.3)	179	17.18***
F/R Luncha *n* (%)	360 (69.1)	153 (81.2)	182 (74.0)	114 (79.2)	809	23.10***
Received Accommodationsb *n* (%)	61 (11.7)	20 (10.7)	30 (12.2)	34 (23.6)	146	15.53***
6th gradec *n* (%)	247 (47.4)	88 (47.1)	135 (54.9)	68 (47.2)	538	4.39
Total Sample *n* (%)e	521 (47.4)	187 (17)	246 (22.4)	144 (13.1)	1,098	-

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

Notes. aF/R Lunch indicates that student qualified for free or reduced lunch in the 15-16 school year. bReceived Accommodations indicates that student had an Individual Education Plan (IEP) or 504 accommodations in the 15-16 school year. c6th grade represents students in 6th grade for 15-16 school year, all other students included in sample were in 7th grade 15-16 school year. χ^2 column indicates chi-square result, *df* = 3. Original Sample includes all students enrolled in intervention schools for Year 1 and Year 2. dPercentage of individual school sample. ePercentage of total sample.

Table 2

Demographic Characteristics of Selected Analysis Sample by School, Year 1

	A	B	C	Total	χ^2
Female-Identified *n* (%)[e]	67 (50.8)	59 (51.3)	34 (55.7)	160	0.89
Male-Identified *n* (%)	65 (49.2)	56 (48.7)	27 (44.3)	148	-
White *n* (%)	14 (10.6)	27 (23.5)	2 (3.3)	43	14.11***
Black *n* (%)	37 (28.0)	17 (14.8)	3 (4.9)	57	15.85***
Hispanic *n* (%)	53 (40.2)	56 (48.7)	46 (75.4)	156	20.81***
Asian *n* (%)	28 (21.2)	14 (12.2)	10 (16.4)	52	3.28
F/R Lunch[a] *n* (%)	91 (68.9)	91 (79.1)	48 (78.7)	231	5.81
Received Accommodations[b] *n* (%)	8 (6.1)	11 (9.6)	4 (6.6)	23	1.30
6th grade[c] *n* (%)	64 (48.5)	57 (49.6)	18 (29.5)	139	47.48*
# Classrooms Year 1	16	9	4 (6.6)	28	-
# Classrooms Year 2	11	8	9 (14.7)	28	-
Total Selected Analysis Sample[d] *n* (%)[f]	132 (42.8)	115 (37.3)	61 (19.8)	308	-

* $p < .05$, ** $p < .01$** $p < .001$***

Notes. [a]F/R Lunch indicates that student qualified for free or reduced lunch in the 15-16 school year. [b]Received Accommodations indicates that student had an Individual Edcuation Plan (IEP) or 504 accommodations in the 15-16 school year. [c]6th grade represents students in 6th grade for 15-16 school year, all other students included in sample were in 7th grade 15-16 school year. [d]Selected Analysis Sample includes all students enrolled in intervention schools for Year 1 and Year 2 that met study inclusion criteria. χ^2 column indicates chi-square result between selected analysis sample and original sample, $df = 2$. [e]Percentage of individual school sample. [f]Percentage of total sample.

Table 3

SDQ and DESSA Time Point Missingness

Measure	*n* with Data (%)	n Missing[a] (%)
Fall Year 1 SDQ[b]	245 (79.3)	64 (20.7)
Spring Year 1 SDQ	252 (81.6)	57 (18.4)
Fall Year 2 SDQ	185 (59.9)	124 (40.1)
Spring Year 2 SDQ	192 (62.1)	117 (37.9)
Fall Year 1 DESSA[c]	274 (88.7)	35 (11.3)
Spring Year 1 DESSA	158 (51.1)	151 (48.9)
Fall Year 2 DESSA	201 (65.0)	108 (35)
Spring Year 2 DESSA	192 (62.1)	117 (37.9)
Selected Analysis Sample[d]	309 (100)	-

Notes. [a]Missing indicates <75% of items completed. [b]SDQ refers to the Strengths and Difficulties Questionnaire (Goodman, Meltzer, & Bailey, 1998), Adapted (11 items). [c]DESSA refers to the Devereux Student Strengths Assessment-Mini, Form A (Naglieri, LeBuffe, & Shapiro, 2009). [d]Selected Analysis Sample includes all students enrolled in intervention schools for Year 1 and Year 2 that met study inclusion criteria.

Table 4

Correlations of SDQ[a] Means, DESSA[b] Total Scores

		M (SD)	1	2	3	4	5	6	7	8
1	Fall 2015 SDQ	0.61 (0.36)	1 (*n*=245)	**.561**** (*n*=200)	**.569**** (*n*=155)	**.524**** (*n*=154)	**-.212**** (*n*=226)	-.161 (*n*=128)	-.142 (*n*=165)	**-.161*** (*n*=154)
2	Spring 2016 SDQ	0.65 (0.40)		1 (*n*=252)	**.645**** (*n*=150)	**.472**** (*n*=154)	-.091 (*n*=224)	-.164 (*n*=128)	.010 (*n*=166)	-.095 (*n*=155)
3	Fall 2016 SDQ	0.56 (0.37)			1 (*n*=185)	**.606**** (*n*=153)	**-.245**** (*n*=165)	**-.212*** (*n*=117)	**-.221**** (*n*=166)	**-.183*** (*n*=123)
4	Spring 2017 SDQ	0.60 (0.37)				1 (*n*=192)	**-.263**** (*n*=171)	**-.205*** (*n*=110)	-.105 (*n*=160)	**-.210*** (*n*=125)
5	Fall 2015 DESSA	22.83 (6.67)					1 (*n*=274)	**.648**** (*n*=149)	**.405**** (*n*=172)	**.441**** (*n*=162)
6	Spring 2016 DESSA	22.23 (7.45)						1 (*n*=158)	**.460**** (*n*=118)	**.415**** (*n*=85)
7	Fall 2016 DESSA	24.76 (6.14)							1 (*n*=201)	**.746**** (*n*=147)
8	Spring 2017 DESSA	24.98 (6.81)								1 (*n*=192)

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

Notes. [a]SDQ refers to the Strengths and Difficulties Questionnaire (Goodman, Meltzer, & Bailey, 1998), Adapted (11 items), respondents answered on 0-2 scale indicating how true (not at all, somewhat, or very) the item was for them. Higher SDQ score denotes more negative (e.g., worse) mental health. [b]DESSA refers to the Devereux Student Strengths Assessment-Mini, Form A (Naglieri, LeBuffe, & Shapiro, 2009). See Appendix D for complete measures.

Table 5

Final Model 1. Predicting Student SEL, Spring Year 1

Predictor		Coefficient Estimate (s.e.)	t value	p
(intercept)	β_{00}	11.43 (3.62)	3.25	0.002**
Year 1 Fall SEL[a]	β_{01}	0.32 (0.07)	4.54	< 0.001***
Year 1 Pro. Implementation Quality[b]	β_{02}	0.21 (1.25)	0.17	0.869
Gender[c]	β_{04}	4.72 (0.96)	4.92	<0.001***
Race (Binary)[d]	β_{05}	3.38 (0.92)	3.69	<0.001***
School A[e]	β_{06}	-4.79 (1.48)	-3.22	0.003**
School C	β_{07}	-4.27 (1.98)	-2.163	0.040*

Variance Components

Level 1	Within Classrooms (Gender)	e	7.82 (S.D.:2.80)
Level 2	Between classrooms (intercept)	r_{0i}	23.46 (S.D.: 4.84)
Residual		r_{1i}	42.33 (S.D.: 6.51)

Goodness of Fit Statistics

AIC: 2088.8
BIC: 2129.8
Deviance: 2066.8

+ p < .10, * p <.05, ** p < .01, *** p < .001

Notes. [a]Measured by DESSA completed by MOSAIC teacher in Fall Year 1 [c]Gender coded: 1 = Female; 0 = Male. [b]Measured by MOSAIC Consultant Rating Year 1. [d]Race (Binary) coded: 1= Asian and White; 0 = All other racial identity (majority Hispanic or Black). [e]Schools dummy coded in reference to School B. The following covariates were included in model building but not included in final model because they were nonsignificant predictors and did not contribute to goodness of fit: MOSAIC Curriculum Condition. *N* students = 308, *n* classrooms = 28.

Table 6

Final Model 2. Predicting Student SEL Spring Year 2

Predictor		Coefficient Estimate (s.e.)	*t* value	*p*
(intercept)	β_{00}	18.41 (3.18)	3.30	<0.001***
Year 2 Fall SEL[a]	β_{01}	0.29 (0.09)	3.30	0.002**
Year 2 Pro. Implementation Quality[b]	β_{02}	-1.25 (1.08)	-1.16	0.255
Gender[c]	β_{04}	1.92 (0.60)	3.21	0.001**
Race (Binary) [d]	β_{05}	1.68 (0.71)	2.38	<0.018*

Variance Components

Level 1	Between students (Year 2 SEL)	e	0.09 (S.D.: .03)
Level 2	Between classrooms (intercept)	r_{0i}	67.53 (S.D.: 67.54)
Residual		r_{1i}	23.32 (S.D.: 4.82)

Goodness of Fit Statistics

   AIC: 1933.3
   BIC: 1966.8
   Deviance: 1915.3

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

  Notes. [a]Measured by DESSA completed by MOSAIC teacher in Fall Year 2 [c]Gender coded: 1 = Female; 0 = Male. [b]Measured by MOSAIC Consultant Rating Year 2. [d]Race (Binary) coded: 1= Asian and White; 0 = All other racial identity (majority Hispanic or Black). The following covariates were included in model building but not included in final model because they were nonsignificant predictors and did not contribute to goodness of fit: MOSAIC Curriculum Condition and school. *N* students = 308, *n* classrooms = 28.

Table 7

Final Model 3. Predicting Negative Mental Health Spring Year 1

Predictor		Coefficient Estimate (s.e.)	*t* value	*p*
Intercept	β_{00}	0.27 (0.11)	2.48	0.019**
Year 1 Fall Neg. Mental Health[a]	β_{01}	0.37 (0.06)	6.08	<0.001***
Year 1 Pro. Implementation Quality[b]	β_{02}	0.03 (0.05)	0.75	0.462
School A[c]	β_{03}	0.18 (0.05)	3.46	0.001**
School C	β_{04}	0.001 (0.07)	0.021	0.984

Variance Components

Level 1	Between students (Y1 Men. He.)	*e*	0.01 (S.D.: .12)
Level 2	Between classrooms (intercept)	r_{0i}	0.00 (S.D.: 0.00)
Residual		r_{1i}	0.12 (S.D.: 0.35)

Goodness of Fit Statistics

    AIC: 259.4
    BIC: 293.0
    Deviance: 241.4

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

Notes. [a]Measured by SDQ completed by student Fall Year 1. [b]Measured by MOSAIC Consultant Rating Year 1. [c]Schools were dummy coded in reference to School B. The following covariates were included in model building but not included in final model because they were nonsignificant predictors and did not contribute to goodness of fit: MOSAIC Curriculum Condition, Race, Gender. *N* students = 308, *n* classrooms = 28.

Table 8

Final Model 4. Predicting Negative Mental Health Spring Year 2

Predictor		Coefficient Estimate (s.e.)	*t* value	*p*
(intercept)	β_{00}	0.85 (0.09)	9.02	<0.001***
Year 2 Fall Neg. Mental Health[a]	β_{01}	0.35 (0.07)	5.18	<0.001***
Year 2 Prog. Implementation Quality[b]	β_{02}	-0.17 (0.04)	-4.76	<0.001***
Race[c]	β_{03}	-.08 (.04)	-2.14	0.034*
School A[d]	β_{04}	-0.13 (0.04)	-2.91	0.005**
School C	β_{05}	0.06 (0.05)	1.25	0.214

Variance Components

Level 1	Within classrooms (Y2 Men. He.)	*e*	0.05 (S.D.: .23)
Level 2	Between classrooms (intercept)	r_{0i}	0.02 (S.D.: 0.13)
Residual		r_{1i}	0.09 (S.D.: 0.30)

Goodness of Fit Statistics

AIC: 168.9
BIC: 202.5
Deviance: 150.9

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

Notes. [a]Measured by SDQ student report Fall Year 2. [b]Measured by MOSAIC Consultant Rating Year 2. [c]Race (Binary) coded: 1= Asian and White; 0 = All other racial identity (majority Hispanic or Black). [d]Schools were dummy coded in reference to School B. The following covariates were included in model building but not included in final model because they were nonsignificant predictors and did not contribute to goodness of fit: MOSAIC Curriculum Condition, Gender. *N* students = 308, *n* classrooms = 28.

Table 9.

Final Model 5. SEL Skill Change as Mediator of Program Implementation predicting Mental Health Year 1a

Predictor		Coefficient Estimate (s.e.)	*t* value	*p*
(intercept)	β_{00}	0.474 (0.107)	4.14	<0.001***
Year 1 Fall Mental Health[b]	β_{01}	0.367 (0.086)	4.29	<0.001***
Year 1 Program Implementation	β_{02}	0.028 (0.044)	0.65	0.524
School	β_{04}	-0.008 (0.003)	-3.08	0.005**
Year 1 SEL Skill Change	β_{05}	-0.012 (0.014)	-0.88	0.378
Year 1 SEL Skill Change*Year 1 Program Implementation	β_{06}	0.003 (0.007)	0.40	0.691

Variance Components

Level 1	Between students (Y1 Men. He.)	e	.01 (S.D.: .07)
Level 2	Between classrooms (intercept)	r_{0i}	0.00 (S.D.: 0.00)
Residual		r_{1i}	0.12 (S.D.: 0.35)

Goodness of Fit Statistics

AIC: 256.3
BIC: 293.6
Deviance: 236.3

* $p <.05$, ** $p < .01$, *** $p < .001$

Notes. aUnstandardised Residuals of DESSA Spring Year 1 Regressed on DESSA Fall Year;. bMeasured by SDQ completed by MOSAIC students in Fall Year 1. The following covariates were included in model building but not included in final model because they were insignificant predictors and did not contribute to goodness of fit: MOSAIC Curriculum Condition, Race, Gender. *N* students = 308, *n* classrooms = 28.

Table 10.

Final Mode 6. SEL Skill Change as Mediator of Program Implementation predicting Mental Health Year 2[a]

Predictor		Coefficient Estimate (s.e.)	*t* value	*p*
(intercept)	β_{00}	0.742 (0.072)	10.324	<0.001***
Year 2 Baseline Mental Health[b]	β_{01}	0.323(0.052)	6.258	<0.001***
Year 2 MOSAIC Consultant Rating	β_{02}	-0.174 (0.038)	-4.63	<0.001***
Race[c]	β_{03}	-0.079 (.040)	-1.99	0.047*
School	β_{04}	0.007 (0.003)	2.66	0.009**
SEL Skill Change Y2[c]	β_{05}	0.002 (0.010)	0.19	0.847
SEL Skill Change Y2[c]* Year 2 MOSAIC Consultant Rating	β_{05}	-0.002 (0.005)	-0.48	0.934

Variance Components

Level 1	Between students (Y2 Men. He.)	*e*	.004 (S.D.: 0.66)
Level 2	Between classrooms (intercept)	r_{0i}	0.09 (S.D.: 0.00)
Residual		r_{1i}	0.10 (S.D.: 0.31)

Goodness of Fit Statistics

    AIC: 183.08
    BIC: 224.11
    Deviance: 161.08

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

Note. [a]Measured by SDQ completed by MOSAIC students in Spring Year 2. [b]Measured by SDQ completed by MOSAIC students in Fall of 2015. Race (Binary) coded: 1= Asian and White; 0 = All other racial identity (majority Hispanic or Black. [c]Unstandardized residual of Spring Y2 SEL skills regressed onto Fall Y2 SEL skills.The following covariates were included in model building but not included in final model because they were insignificant predictors and did not contribute to goodness of fit: MOSAIC Curriculum Condition, Gender. *N* students = 308, *n* classrooms = 28.

Table 11

Final Model 7. Fall Year 1 SEL* Year 1 Program Implementation Quality as Predictor of Negative Mental Health Spring Year 1

Predictor		Coefficient Estimate (s.e.)	*t* value	*p*
(Intercept)	β_{00}	0.04 (0.32)	0.13	0.90
Year 1 Fall SEL[a]	β_{01}	0.02 (0.01)	1.23	0.22
Year 1 Pro. Implementation Quality[b]	β_{02}	0.28 (0.14)	2.03	0.04*
Gender[c]	β_{03}	0.09 (0.04)	2.00	0.05*
School A[d]	β_{04}	0.13 (0.05)	2.80	0.005**
School C	β_{05}	-0.03 (0.06)	-0.52	0.60
Year 1 Fall SEL* Year 1 Pro. Implementation Quality	β_{06}	-0.01(0.01)	-1.60	0.11†
Variance Components				
Level 2 Between classrooms (intercept)	r_{0i}	0.00 (S.D.: 0.00)		
Residual	r_{1i}	0.14 (S.D.: 0.38)		
Goodness of Fit Statistics				
AIC: 292.3				
BIC: 325.9				
Deviance: 274.3				

† *p * p <.11, * p <.05, ** p < .01, *** p < .001*

Notes. [a]Measured by DESSA teacher rating Fall Year 1. [b]Measured by MOSAIC Consultant Rating Year 1. [c]Gender coded: 1 = Female; 0 = Male. [d]Schools were dummy coded in reference to School B. The following covariates were included in model building but not included in final model because they were nonsignificant predictors and did not contribute to goodness of fit: MOSAIC Curriculum Condition, Race. *N* students = 308*, n* classrooms = 28.

Table 12

Final Model 8. Fall Year 2 SEL* Year 2 Program Implementation Quality as Predictor of Negative Mental Health Spring Year 2

Predictor		Coefficient Estimate (s.e.)	t value	p
(Intercept)	β_{00}	0.52 (0.27)	1.96	0.05
Year 2 Fall SEL	β_{01}	0.08 (0.12)	2.25	0.03*
Year 2 Prog. Implementation Quality	β_{02}	0.08 (0.12)	0.70	0.48
Race	β_{03}	-0.12 (0.04)	-2.98	0.003**
School A	β_{04}	-0.21 (0.06)	-3.79	0.001**
School C	β_{05}	0.03 (0.06)	0.57	0.58
Year 2 Fall SEL* Year 2 Program Implementation Quality	β_{06}	-0.01 (0.005)	-2.31	0.02*
Variance Components				
Level 2 Between classrooms (intercept)	r_{0i}	0.00 (S.D.: 0.03)		
Residual	r_{1i}	0.11 (S.D.: 0.33)		
Goodness of Fit Statistics				

AIC: 214.1
BIC: 247.7
Deviance: 196.1

* $p <.05$, ** $p < .01$, *** $p < .001$

Notes. [a]Measured by DESSA teacher rating Fall Year 2. [b]Measured by MOSAIC Consultant Rating Year 2. [c]Race (Binary) coded: 1= Asian and White; 0 = All other racial identity (majority Hispanic or Black). [d]Schools were dummy coded in reference to School B. The following covariates were included in model building but not included in final model because they were nonsignificant predictors and did not contribute to goodness of fit: MOSAIC Curriculum Condition, Gender. *N* students = 308*, n* classrooms = 28.
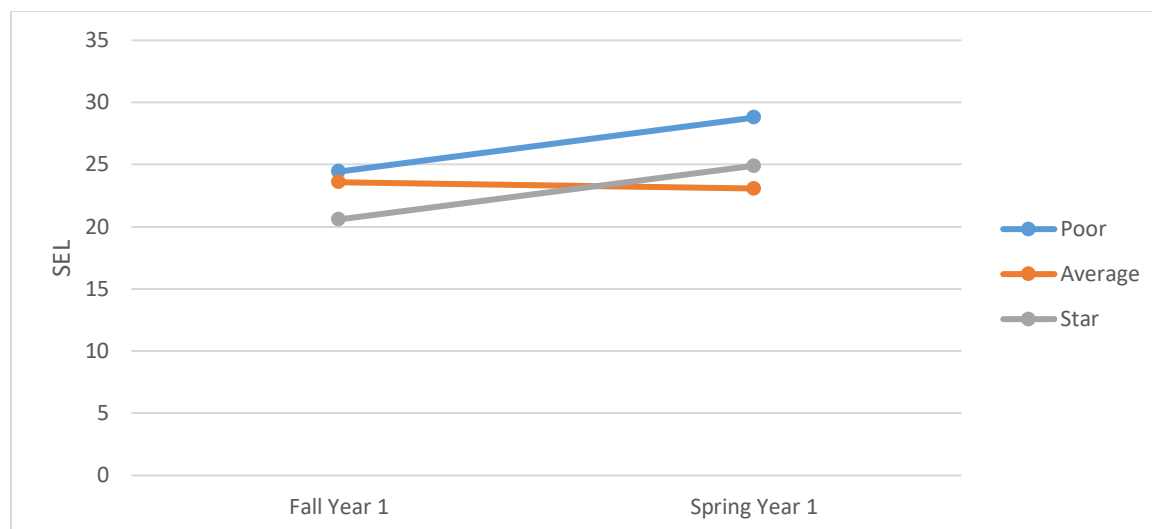
Figure 1. Student SEL (DESSA) from Fall to Spring Year 1, by Year 1 program implementation
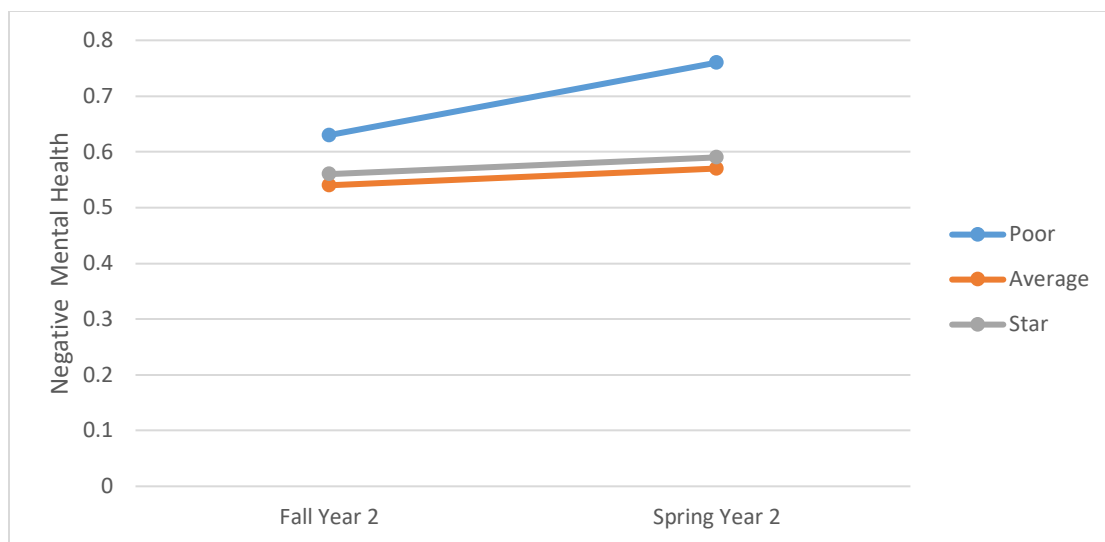quality (MOSAIC Consultant Rating of poor, average, star)
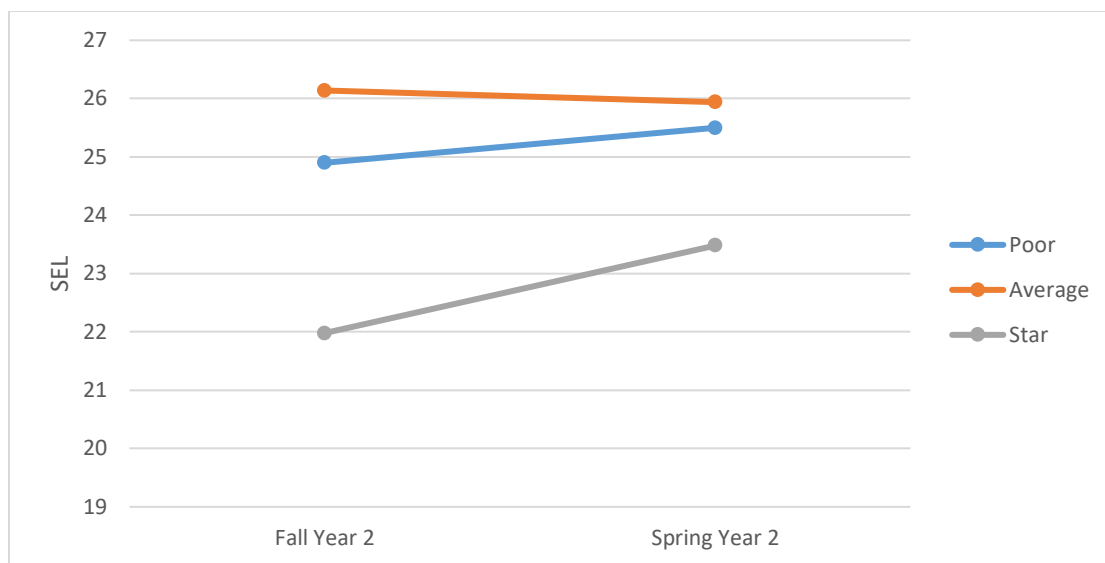
Figure 2. Student SEL (DESSA), from Fall to Spring Year 2, by Year 2 program implementation quality (MOSAIC Consultant Rating of poor, average, star)
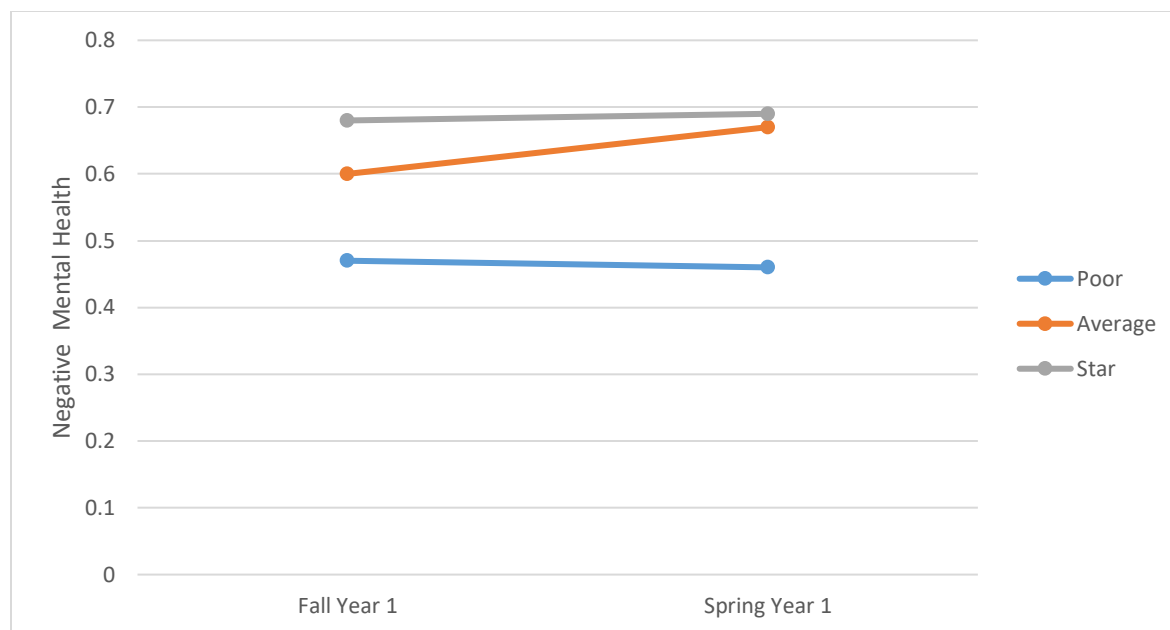
Figure 3. Student negative mental health (SDQ), from Fall to Spring Year 1, by program implementation quality (MOSAIC Consultant Rating of poor, average, star)
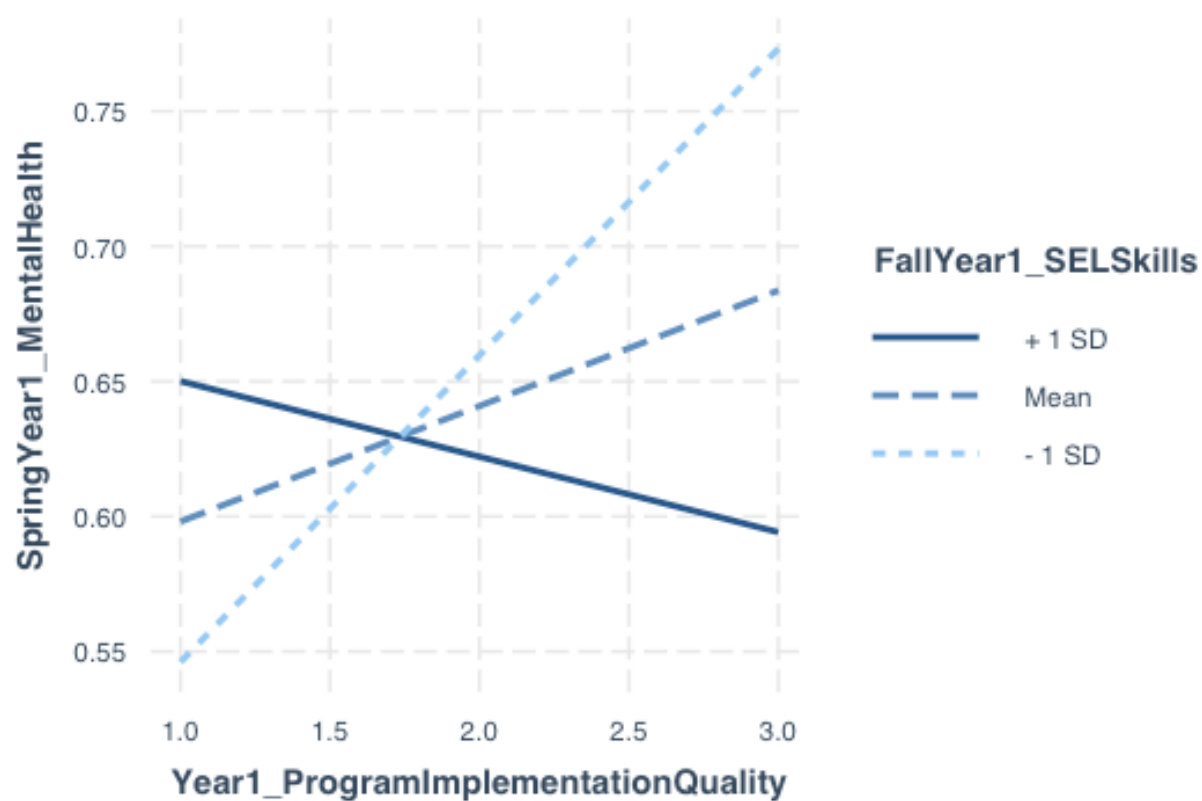
Figure 4. Student negative mental health (SDQ) from Fall to Spring Year 2, by Year 2 program implementation quality (MOSAIC Consultant Rating of poor, average, star)

Figure 5. Interaction plot of Fall Year 1 SEL (DESSA) and Year 1 program implementation quality (MOSAIC Consultant Rating) as predictor of Spring Year 1 negative mental health (SDQ)
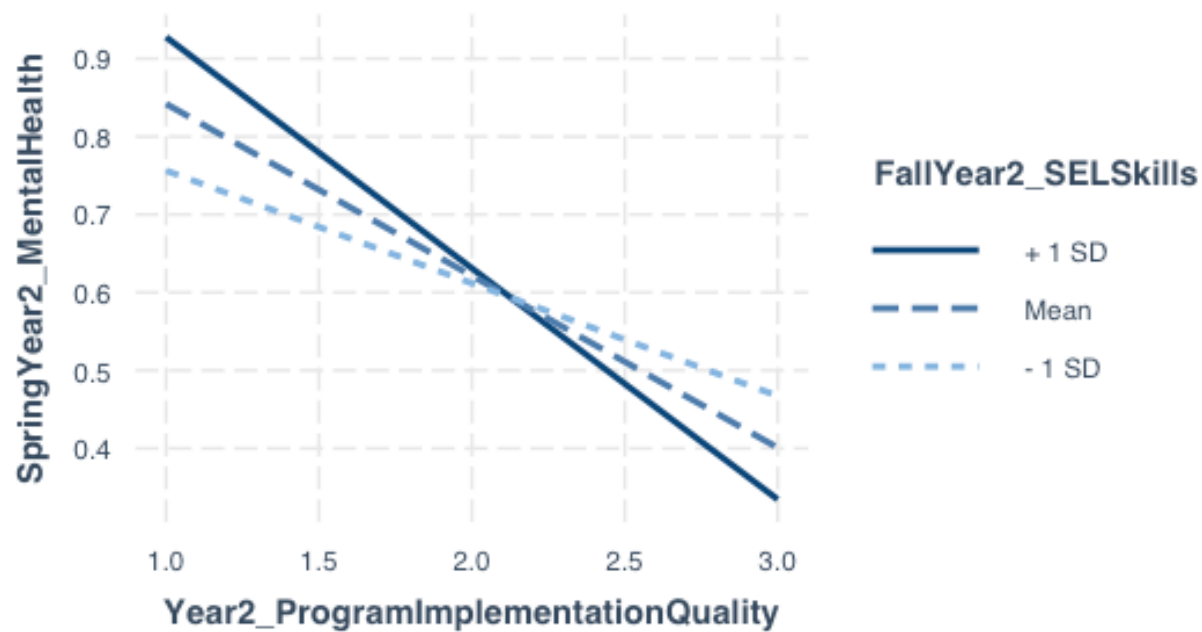
Figure 6. Interaction plot of Fall Year 2 SEL (DESSA) and Year 2 program implementation quality (MOSAIC Consultant Rating) as predictor of Spring Year 2 negative mental health (SDQ)

List of Appendices

Appendix A

MOSAIC Virtues & Skills

Positive Purpose

Supporting Virtues
1) Constructive Creativity
2) Helpful Generosity
3) Optimistic Future-Mindedness
4) Responsible Diligence
5) Compassionate Forgiveness
   and Gratitude

Supporting Skills
1) Emotional Regulation
2) Communication
3) Empathy
4) Social Problem Solving

MOSAIC VIRTUES & SKILLS BY MONTH

MONTH	THEME	VIRTUE	SKILLS
September	*Why are we here: Finding Our Positive Purpose*	Introduction to Positive Purpose	Communication & Social Problem Solving
October	*What Kind of Person Do I Want to Be?*	Virtue Identification	Skill Identification
November	*Making Ourselves, School, and World Better*	Constructive Creativity	Communication & Social Problem Solving
December	*Giving Back to Ourselves, School, and World*	Helpful Generosity	Communication & Social Problem Solving
January	*Planning for the Future*	Optimistic Future-Mindedness	Empathy & Social Problem Solving
February	*Showing Resilience and Overcoming Obstacles*	Responsible Diligence	Emotion Regulation & Social Problem Solving
March	*Appreciating Ourselves, Our School, and the World*	Compassionate Gratitude	Communication & Empathy
April	*Connecting with Others and Being a Leader*	Compassionate Forgiveness	Emotion Regulation & Empathy
May	*Looking Forward: Next Steps on the Journey*	Positive Purpose	Communication & Social Problem Solving
June	*Looking Back: What Have I Accomplished? What Have I Learned?*	All Virtues Summary	All Skills Integrated

Appendix B.

Data Collection Map

Construct *(Measure)*	Level	Classroom Year 1		Classroom Year 2	
		Fall Year 1 *(2015)*	Spring Year 1 *(2016)*	Fall Year 2 *(2016)*	Spring Year 2 *(2017)*
Program Implementation Quality (MOSAIC Consultant Rating)	Level 2 (classroom)		X		X
Social-Emotional Learning (SEL) (DESSAb- teacher report)	Level 1 (student)	X	X	X	X
Negative Mental Health (SDQc- self-report)	Level 1 (student)	X	X	X	X

Notes. aLevel = Multilevel Model Level. b Devereux Student Strengths Assessment-Mini, Form A (Naglieri, LeBuffe, & Shapiro, 2009). c Strengths and Difficulties Questionnaire (Goodman, Meltzer, & Bailey, 1998), Adapted (11 items).

Appendix C
MOSAIC Consultant School Visit Checklist

Instructions: This checklist is intended to serve 2 purposes: (1) as a guide outlining the components of a typical school visit and (2) to document and track information about these components. Note that we are asking for 3 takeaways as a guideline and as a way to keep us focused; you can include fewer or more as necessary.

General Information:

- Date:
- School Name:
- Consultant name:

- **To Accomplish for Today (in order of priority):**

- **NEW Action Steps from today (please asterisk * any continuing steps):**

Ambassador Attendance:

- 6th-7th Grade Ambassador STAT Training Attendance [DATE]:

- 8th Grade Ambassador Positive Purpose Project Team Meeting Attendance [DATE]:

MOSAIC Individual Meetings

- Did you have any meetings/conversations with School Administrator?
    - Yes
    - No

3 take-aways:

- Did you have a meeting/conversation with the **Instructional Support Team Liaison?**
    - Yes
    - No

3 take-aways:

- Did you have a meeting/conversation with the **STAT Point Person?**
    - Yes
    - No

3 take-aways:

- Did you have a meeting/conversation with the **<u>Positive Purpose Project Point Person?</u>**
  - Yes
  - No

3 take-aways:

MOSAIC Group Meetings:

- Did you attend an <u>Instructional Support Team meeting OR learn that one took place since your last visit?</u>
  - Yes
  - No

3 take-aways:

- Did you attend a <u>STAT (6th and 7th grade) Ambassador Training OR learn that one took place since your last visit?</u>
  - Yes
  - No

3 take-aways:

- Did you attend a <u>Positive Purpose Project (8th grade) Ambassador Meeting OR learn that one took place since your last visit?</u>
  - Yes
  - No

3 take-aways:

General Classroom Observations (MOSAIC or non-MOSAIC)

It is <u>not</u> expected that you will visit classrooms every time you are in the school.

- Number of classrooms visited/observed in total: __
- <u>What type of classrooms/lessons? (MOSAIC/non-MOSAIC)</u>
  - If MOSAIC (if you visited whole lesson or just saw a quick glimpse), please comment on whether classrooms seemed to be doing MOSAIC and list teacher names:
  - If non-MOSAIC, please comment on use of MOSAIC skills/virtues or Throughline in the lesson:

School-wide observations (major events, school climate, etc):

3 take-aways:
Additional Notes:

Appendix D
Measures

Devereux Student Strengths Assessment-Mini, Form A (Naglieri, LeBuffe, & Shapiro, 2009)

	Never	Rarely	Occasionally	Frequently	Very Frequently
Accept responsibility for what she/he did	0	1	2	3	4
Do something nice for somebody?	0	1	2	3	4
Speak about positive things?	0	1	2	3	4
Pay attention?	0	1	2	3	4
Contribute to group efforts?	0	1	2	3	4
Perform the steps of a task in order?	0	1	2	3	4
Show care when doing a project or school-work?	0	1	2	3	4
Follow the advice of a trusted adult?	0	1	2	3	4

Strengths Difficulties Questionnaire- Selected Items (Goodman et al., 1998)

Item (Subscale)	Not True	Somewhat True	Certainly True
I get a lot of headaches, stomach-aches or sickness. (Emotional Problems)	0	1	2
I worry a lot. (Emotional Problems)	0	1	2
I have many fears, I am easily scared. (Emotional Problems)	0	1	2
I am nervous in new situations. I easily lose confidence.(Emotional Problems)	0	1	2
I get very angry and often lose my temper. (Conduct Problems)	0	1	2
I fight a lot. I can make other people do what I want. (Conduct Problems)	0	1	2
I am often accused of lying or cheating. (Conduct Problems)	0	1	2
I would rather be alone than with people my age. (Peer Problems)	0	1	2
I have one good friend or more. (Peer Problems)	0	1	2
I get along better with adults than with people my own age. (Peer Problems)	0	1	2
Other children or young people pick on me or bully me. (Peer Problems)	0	1	2

Appendix E
Missing Data Figure

1,098 Students Enrolled in Study Schools

Remaining *n* = 1,002 students

Deleted *n = 96*
Missing SDQ at all of the study time points (*n* = 57) or DESSA at all (*n* = 39),

Remaining *n* = 842 students

Deleted *n* = 160
Missing MOSAIC Consultant Rating in 2015-16 (n = 156) or 2016-17 (n = 4)

Remaining *n* = 426

Deleted *n* = 416
MOSAIC Consultant Rating "not enough information" 2015-16 ((*n* = 276) or 2016-17 classroom (*n* = 140)

Reminaing *n* = 317

Deleted *n* = 109
Invalid SDQ Response at any time point

Remaining *n* = 309

Deletd *n* = 8
Remaining students from School D

Remaining *n* = 308

Deleted *n* = 1
Participant had invalid SDQ score. Once score deleted student no longer had valid SDQ score at any time point.

Subsample A
2016 Student Activity Rating
Remaining *n* =

Deleted *n* = 188(missing)
Deleted n = 57 (Invalid)

Subsample B
2017 Student Activity Rating
Remaining *n* = 100

Deleted *n* = 169 (missing)
Deleted *n* = 39 (invalid)