# Changes in Classroom Quality Predict Ghanaian Preschoolers' Gains in Academic and Social-Emotional Skills

Dana Charles McCoy, *Harvard Graduate School of Education*

Sharon Wolf, *University of Pennsylvania Graduate School of Education*

Address all correspondence to first author at: dana_mccoy@gse.harvard.edu; Larsen Hall Room 704, 14 Appian Way, Cambridge, MA 02138, USA

In press, *Developmental Psychology*

**Acknowledgements:** We thank the UBS Optimus Foundation and the World Bank Strategic Impact Evaluation Fund (SIEF) for direct support for this project. In addition, we thank the NYU Abu Dhabi Research Institute for core support to the Global TIES for Children Center, which provides intramural support to the second author for research activities that directly contributed

to this paper.

**Abstract**

Rates of participation in early childhood education (ECE) programs are on the rise globally, including in sub-Saharan Africa. Yet little evidence exists on the quality of these programs and on the role of classroom quality in predicting learning for young children across diverse contexts. This study uses data from the Greater Accra Region of Ghana ($N$ = 3,407; $M_{age}$ = 5.8 years; 49.5% female) to examine how changes in four culturally-validated dimensions of ECE classroom quality predict children's growth in early academic and social-emotional skills from the beginning to the end of one academic year. We find that improvements in domains of classroom instructional quality are related to small, positive gains in children's early academic and social-emotional outcomes over the school year, and that these improvements are generally larger for children and classrooms with higher baseline proficiency and quality levels. Associations between changes in social-emotional aspects of classroom quality and child outcomes were mixed. These results extend the knowledge base on ECE quality to a new and under-represented context while also providing important information regarding the contexts and children for whom teacher training and other quality-focused improvement efforts may be most needed.

**Key words**: early childhood development, early childhood education, classroom quality, Ghana

# Changes in Classroom Quality Predict Ghanaian Preschoolers' Gains in Academic and Social-Emotional Skills

There are vast inequalities in children's developmental opportunities and outcomes around the world (Black et al., 2017; McCoy et al., 2016). Compared to other regions, sub-Saharan Africa (SSA) has the largest number of young children experiencing malnutrition and poverty (Black et al., 2017), as well as the largest number and proportion of three- and four-year-olds (29.4 million; 44%) failing to meet cognitive and social-emotional milestones (McCoy et al., 2016). Global initiatives such as the Sustainable Development Goals aim to reduce these inequalities by expanding access to learning opportunities, with Target 4.2 seeking to "ensure that all girls and boys have access to quality early childhood development, care and pre-primary education" by 2030 (United Nations, 2015). At the same time, little evidence exists on how to measure the quality of early learning programs or their effects on children's outcomes in SSA.

Correlational research in the United States has found positive associations between the quality of early childhood education (ECE) environments and child outcomes (e.g., Curby et al., 2009; Hamre, Hatfield, Pianta, & Jamil, 2014; Mashburn et al., 2008; Rimm-Kaufman et al., 2009). Children who attend preschools with high levels of warmth, organization, and instructional support tend to show greater average gains in both early academic skills and social-emotional wellbeing relative to their peers in lower quality environments. At the same time, research is mixed with regard to the consistency of these results across diverse settings, outcomes, and study designs, with many high-quality studies and meta-analyses showing weak – or even null – associations between classroom quality and child outcomes (Auger, Farkas, Burchinal, Duncan, & Vandell, 2014; Gonzalez, McCoy, & Sabol, 2017; Gordon et al., 2017; Keys et al., 2013). Furthermore, despite evidence demonstrating the evolving nature of classroom processes over time (e.g., Kuger, Kluczniok, Kaplan, & Rossbach, 2016), most ECE research has focused on classroom quality as static, rather than as a set of dynamic processes that may change over time. Finally, almost no research has explored how ECE quality predicts child outcomes in SSA, where developmental challenges are prevalent and access to programming is rapidly expanding (Britto, Yoshikawa, & Boller, 2011).

The aim of the present study is to examine several observed, culturally-validated dimensions of ECE classroom quality as predictors of Ghanaian children's growth in early academic and social-emotional skills from the beginning to the end of one academic year. In particular, we examine whether changes (i.e., improvements) in classroom quality over one school year predict gains in children's outcomes in the same year when accounting for classrooms' starting levels of quality at the beginning of the school year. We also probe the consistency of these findings by examining whether changes in classroom quality are more or less beneficial in supporting children's development based on (1) classrooms' beginning-of-year (fall) levels of quality and (2) children's beginning-of-year (fall) levels of developmental skill. In doing so, we aim to extend the research base on ECE quality to an under-represented context while also considering the contexts for which and children for whom quality-focused improvement efforts may be most beneficial.

**Early Childhood Classroom Quality and Child Outcomes**

Around the world, ECE participation is on the rise. In the United States, enrollment in publicly funded preschool has doubled over the past decade, with states currently serving nearly 1.5 million three- and four-year olds, or approximately one-third of children in this age group (Barnett et al., 2017). In Europe, several countries – including Spain, France, and Belgium – have already achieved universal access to ECE for children three to five (OECD, 2015). Access to ECE is somewhat lower in low- and middle-income country (LMIC) regions, ranging from an average of 17.9 percent of three- and four-year-olds in sub-Saharan Africa to 61.7 percent in Latin America and the Caribbean (McCoy et al., 2017). Despite clear evidence that more children are participating in ECE, research on the quality of these programs is less established.

A number of studies have shown that measures of process quality – including the ways that teachers interact with and respond to children within the classroom setting – are positively associated with modest improvements in children's math, literacy, and behavioral development (Curby et al., 2009; Howes et al., 2003; 2008; Mashburn et al., 2008; Leyva et al., 2015; Vandell, 2004). These measures of process quality focus not only on instruction in the classroom setting, but also on the ways that teachers organize and structure class time, manage children's behavior, and respond to children's social and emotional needs (Pianta et al., 2005).

Importantly, evidence also suggests that the benefits of classroom process quality for children's outcomes are not necessarily uniform (Zaslow et al., 2016). Several studies have shown, for example, that instructional quality is more strongly related to children's early academic outcomes, whereas measures of emotional support are more strongly predictive of children's social, emotional, and behavioral skills (Burchinal, Kainz, & Cai, 2011; Curby et al., 2009; Mashburn et al., 2008). Furthermore, a growing body of research suggests there may be "thresholds" of classroom quality at which greater benefits accrue (Burchinal et al., 2016; Burchinal, Vandergrift, Pianta, & Mashburn, 2010; Hatfield, Burchinal, Pianta, & Sideris, 2016; Weiland, Ulvestad, Sachs, & Yoshikawa, 2013). This work shows that the quality of teacher-child interactions and instruction are more strongly related to children's social-emotional and early academic skills, respectively, in classrooms with higher levels of quality. Finally, the benefits of positive classroom quality may be greatest for children facing low levels of initial skill. For example, Hamre and Pianta (2005) found that increments in classroom emotional support were only beneficial for children facing initially high levels of "functional risk," as defined by low levels of attention and social skills, and/or high levels of behavior problems. Additional research probing the differential contributions of classroom quality is needed to confirm and extend this work.

**Limitations of Existing ECE Quality Research**

There are several limitations of the existing evidence base on ECE quality that are worth noting. First is a set of methodological challenges, including difficulties with measurement of classroom quality and establishing causal inference. Analyses using the Classroom Assessment Scoring System (CLASS), a popular measure of classroom quality developed in the United States, have shown inconsistencies in its reliability and factor structure across studies (Burchinal, 2017; Mashburn, 2017; Sandilos, DiPerna, & Family Life Project Key Investigators, 2014). In addition, a reliance on cross-sectional, correlational data has limited the field's ability to make causal claims regarding the relation between quality and child outcomes. Indeed, studies that attempt to account for selection bias (e.g., through including large numbers of covariates or using quasi-experimental designs such as instrumental variables analysis) often find results that are attenuated compared to those that do not (e.g., Auger et al., 2014; Gonzalez, McCoy, & Sabol, 2017; Yoshikawa et al., 2015).

Second, despite a relatively large body of evidence examining cross-sectional associations between levels of classroom quality and child outcomes, relatively little work has examined how *changes* – increases or decreases – in classroom quality may predict children's development over time. Dynamic systems theory and bioecological models have long highlighted the evolving nature of individuals and settings (Bronfenbrenner & Morris, 2006;

Thelen & Smith, 1996). Indeed, research has shown that classrooms are not static entities (Kuger et al., 2016). Relationships between students and teachers evolve over the course of a school year, and teachers – particularly those participating in professional development and other quality-improvement programs – will often seek out ways to improve their practice. Despite the dynamic nature of classrooms, relatively little work has considered how changes in classroom quality may predict learning over and above initial levels of quality. One notable exception is work from the NICHD Study of Early Child Care, which found that initial levels of quality at six months were positively predictive of both early academic and social-emotional outcomes for children, whereas improvements in quality across five points between six and 54 months were only related to early academic skills (NICHD, 2002).

A third limitation of the collective body of work on ECE quality is its near-exclusive focus on the U.S. and other high-income countries. Challenges with conceptualizing and measuring process quality in culturally diverse and often under-resourced LMICs have limited research in these settings (Myers, 2004). As a result, relatively little is known about what process quality looks like and how it relates to children's outcomes in diverse parts of the world. Several important exceptions do exist, and their findings largely – though not universally – support positive associations between process quality and children's learning. In a study of a preschool program in rural Bangladesh, Aboud (2006) found that classroom quality rated on the Early Childhood Environment Rating Scale-Revised (ECERS-R) was positively correlated with group-level verbal and nonverbal reasoning scores. Several studies have also examined associations between ECE quality and child outcomes in Chile, which prior to 2012 was classified as a middle-income country. Herrera and colleagues (2005) found positive and persistent associations between the ECERS and Chilean children's vocabulary, social skills, and adaptive behavior.

More recently, Leyva and colleagues (2015) found positive associations between the instructional quality dimension of the CLASS and the early writing and executive function skills of children in 91 Chilean preschool classrooms. Using thresholds similar to those identified in the study by Burchinal and colleagues (2010), these authors also found positive associations between classroom organization and children's language, writing, and numeracy skills within high-quality classrooms only (Leyva et al., 2015). At the same time, Yoshikawa and colleagues (2015) showed positive impacts of a teacher workshop and coaching intervention on classroom outcomes that did not translate into gains for child outcomes for Chilean children. In kindergarten classrooms in Ecuador, however, a study in which children were randomly assigned to teachers across 204 schools found that a one standard deviation increment in total CLASS scores predicted increments in children's language, math, and executive function of 0.11, 0.11, and 0.07 standard deviations, respectively (Araujo et al., 2016). Finally, a recent study in rural Indonesia found significant and positive cross-sectional links between observed classroom process (but not structural) quality from the ECERS and children's development (Brinkman et al., 2017). To our knowledge, no research to date has explicitly examined the association between classroom process quality and child outcomes in SSA.

**Early Childhood Education in Ghana**

In the present paper, we focus on the ECE quality received by a sample of children living in the Greater Accra Region of Ghana, the country's second most populous and fastest growing region. Ghana is a lower-middle-income country in West Africa that gained its independence from Great Britain in 1957. Although life expectancy, gross domestic product (GDP), and school enrollment rates have risen dramatically in Ghana over the past several decades (World Bank, 2017), poverty rates remain high, with 20.6 percent of adults and 28.3 percent of children currently living under the national poverty line (Cooke, Hague, & McKay, 2016).

Ghana is considered a leader in Africa with regard to its ECE policy and access levels. In 2007, two years of pre-primary (i.e., kindergarten or "KG") education was introduced as part of the country's education reforms, making it part of the free, compulsory, and universal basic education system. Recent estimates from 2015-2016 place Ghana's pre-primary net enrollment rates at 80 percent (Ghana Ministry of Education, 2016), which is nearly four times higher than average for SSA (21.9%; UNESCO, 2015). In the Greater Accra Region, ECE enrollment rates are as high as 94 percent in some communities (Bidwell, Perry, & Watine, 2014).

Ghana's 2004 National Early Childhood Care and Development Policy highlights access to quality kindergarten education as central to fostering early learning and preventing developmental delays in contexts of adversity. Both private schools – which serve the majority of KG students – and public schools are required to follow the national Ghana Education System (GES) curriculum, which emphasizes play-based, child-centered approaches to supporting children's holistic learning. However, a 2012 report by the Ministry of Education concluded that KG educational quality was low and the curriculum rarely implemented (GES, 2012). A recent study suggests that Ghanaian teacher educators struggle to generalize pedagogical principles, instead focusing on routinized implementation of specific practices and materials during pre-service training (Akyeampong, 2017). These quality-related challenges may help to explain why, despite high ECE enrollment, approximately one-third of Ghanaian preschool-aged children struggle to meet basic developmental milestones such as following directions, working independently, and getting along with others (McCoy et al., 2016).

In the present study, we aim to explore this possibility explicitly in this unique cultural and policy context. As highlighted above, children in Ghana face a unique combination of environmental risks (e.g., poverty, low at-home parental stimulation) and opportunities (e.g., high rates of ECE participation). Bioecological systems theory suggests that children's learning and development is inextricably linked with these environmental characteristics, as well as broader cultural expectations and norms (Bronfenbrenner & Morris, 2006). Indeed, ethnographic research from cross-cultural psychology (Super, Harkness, Barry, & Zeitlin, 2011) and population analyses (Bornstein et al., 2015) have shown that differences in parenting practices in sub-Saharan African may be responsible for unique patterns of development early in life. At the same time, these developmental theories and lessons have not been applied in educational research, leading to open questions with regard to whether and how culturally-specific dimensions of classroom quality might relate to children's outcomes over time.

One of the challenges with measuring – and, by extension, improving – quality in Ghana and other LMICs is the lack of culturally sensitive instruments available for understating process quality in these settings. In the present study, we rely on an instrument called the Teacher Instructional Practices and Processes System (TIPPS ©; Seidman, Raza, Kim, & McCoy, 2013) to assess three distinct, observable dimensions of process quality designed for LMICs (see Wolf, Raza et al., 2018 for further details). First, we focus on how Ghanaian ECE teachers *facilitate deeper learning* by providing feedback, scaffolding learning, and connecting lessons to teaching objectives. To complement this instructionally-oriented dimension of quality, we also assess teachers' *emotional support and behavior management* – including their use of positive and consistent reinforcement strategies – as well as the degree which they *support student expression* by using students' ideas and interests to inform class activities, encourage reasoning and problem solving, and draw connections between subject matter and students' daily lives. Along with these observed process quality dimensions, we review the extent to which teachers use *developmentally appropriate pedagogical practices* as specified in the national curriculum (GES, 2004). Together, these four dimensions reflect core dimensions of process quality identified for use in international settings (e.g., Meyers, 2004; UNESCO, UNICEF, Brookings Institution, & the World Bank, 2017).

**The Present Study**

The objective of the present study is to unpack the relations between changes in early childhood classroom quality and gains in Ghanaian children's early academic and social-emotional skills from the beginning to the end of one school year. To meet this objective, we use data from a randomized experimental evaluation of a teacher training and parental awareness intervention conducted in peri-urban Ghana (Wolf, Aber & Behrman, 2018). The original evaluation found that the teacher training impacted (1) teachers' professional well-being, including substantially reducing teacher burnout and turnover mid-year; (2) classroom quality, including emotional support and behavior management, support for student expression, and curriculum checklist quality; and (3) children's developmental outcomes, including their social-emotional, early literacy, and early numeracy skills. At the same time, the original evaluation assessed only direct treatment impacts and did not examine relations amongst the outcomes themselves.

In the present study, we build on these results to address two primary research aims. First, we attempt to narrow in on the specific dimensions of classroom quality that may matter most for various child outcomes, independent of treatment status. To do so, we examine the links between changes in four different dimensions of classroom quality relevant to the Ghanaian context and gains in children's early academic outcomes (including literacy and numeracy skills) and social-emotional skills (including prosocial skills, executive function, and approaches to learning). We hypothesize that improvements in classroom quality be positively related to gains in children's outcomes at the end of the school year, above and beyond the initial levels of quality within the classroom at the beginning of the school year. We hypothesize stronger associations between improvements in instruction-oriented dimensions of quality (facilitating deeper learning, delivery of curricular content) and gains in early academic outcomes, as well as between improvements in social-emotional dimensions of quality (emotional support and behavior management, supporting student expression) and gains in children's social-emotional outcomes. Our second aim is to further unpack these processes by examining whether improvements in classroom quality across one school year have differential benefits for children depending on levels of (1) classroom quality and (2) children's skills at the beginning of the school year. Consistent with a compensatory model (Miller, Farkas, Vandell, & Duncan, 2014), we hypothesize that gains in quality will be particularly beneficial for children from classrooms with low initial quality and with low levels of baseline skills.

In addressing these aims, the present study contributes to intervention and policy efforts designed to expand access to high-quality ECE programming for children living in Ghana and other LMICs. In particular, we aim to provide a more nuanced understanding of if and how different dimensions of classroom quality are beneficial for different child outcomes, using culturally-validated observational measures and longitudinal data. This study also serves as the largest and most in-depth study of ECE classroom quality in SSA to date. As such, our objective is to expand a robust body of literature from high-income countries to a population of children who stand ready to benefit from ECE services, and to inform teacher professional development efforts as countries across the SSA region expand their ECE systems.

## Methods

**Participants & Procedures**

Data for the present study come from the first year of a school-randomized impact evaluation of a teacher professional development and parent engagement program conducted in the Greater Accra Region, Ghana (Wolf et al., 2018). Schools were randomized to one of three treatment arms: (1) teacher training (82 schools, 153 teachers), (2) teacher training plus parental-awareness training (79 schools, 148 teachers), and (3) control group (79 schools, 143 teachers). This study received Institutional Review Board approval from Innovations for Poverty Action (#1328; “Quality Preschool for Ghana”), University of Pennsylvania (#825679; “Quality Preschool for Ghana”), and New York University (#FY2015-10; “An Intervention to Improve Preschool Quality in Peri-urban Ghana”). A total of 108 public and 132 private schools were selected from six of the nine most disadvantaged districts in the region during the summer of 2015 using a stratified design by district and education sector (public and private). Baseline equivalency across treatment conditions was established for schools, teacher/classroom, and child characteristics (Wolf et al., 2018).

Within schools, passive consent of all Kindergarten 1 (KG1) and Kindergarten 2 (KG2) children was sought via a form sent home to caregivers. Ten caregivers refused their children’s participation. Of the rest, 15 children (8 from KG1, *M* age = 5.3 years in the spring; and 7 from KG2, *M* age = 6.4 years) were randomly selected from each school roster to participate in direct assessments. A small percentage of schools (10%) had one combined KG classroom rather than two separated by level. In these schools, 15 children were randomly selected from the combined class. In schools with fewer than 15 KG children, all children were sampled.

Trained, multi-lingual assessors implemented direct assessments to children in the fall (September-October, 2015) and in the spring (May-June, 2016). Assessments were conducted in school and took place in a quiet setting away from the normal classroom environment. Assessors spent several minutes chatting and playing games with children to make them comfortable before beginning the assessment. As schools in this sample reported using a mixture of English and local language for instruction, part of this initial introduction was intended to help the assessor to gauge children's linguistic preferences. Assessors then administered the assessment in the language he/she deemed most appropriate for the child, including: Twi/Fanti only (39.0%), Ewe only (1.3%), Ga only (5.0%), English only (37.9%), and mixed English and local language (16.9%). All assessors had prior experience working with children, received five days of training by a certified Master Trainer, and participated in two days of field practice following protocols developed by the child development tool developers (Pisani, Dowd, & Borisova, 2015). Inter-rater reliability was established during practice sessions, with kappa values averaging .86 across domains (*range* = .67-.97). Data were also collected on child and household demographics from primary caregivers (41.6% mothers, 44.6% fathers, 13.8% other) via a telephone survey at the beginning of the school year, and/or from school records.

In both the fall and the spring, teachers were videotaped teaching a lesson in their classrooms for 30-60 minutes of the approximately six-hour school day (see Mashburn et al., 2014 and Seidman et al., 2017 for a discussion on conducting video-based teacher assessments). The first five minutes of the lesson in the video was viewed passively by a trained rater, and then the next 25 minutes were coded using two instruments: a curriculum pedagogical practice checklist, and an observational tool to assess the quality of the classroom environment. All raters were Ghanaian nationals and had at least a Bachelor's degree. Raters attended a five-day training that included a review of the measurement protocols, bias awareness and minimization strategies, and practice coding using real classroom videos. Upon completion of the training, each rater had to meet specific calibration criteria within three attempts prior to coding study videos.

**Child sample.** In total, 3,435 children – an average of 14.3 (*range* = 4-15) per school – were selected for participation at baseline (fall). At follow-up (spring), 2,975 (86.6%) of those children remained in their schools. At follow-up, 432 children were added to the sample based on two scenarios: (1) to replace children who had left the school with children from the existing classroom roster, and (2) to add children to schools that had fewer than 15 KG children in the fall where enrollment increased by the spring. The final analytic sample for the present study includes the 3,407 children with available spring data. Approximately half of children (50.45%) were male, and the average age in the spring was 5.82 years ($SD = 1.33$).

Children who entered the sample in the spring did not differ from those who entered in the fall by age or gender. However, they had lower scores on three spring outcomes: executive function (45.0 vs. 47.2%, $t = 2.07$, $p < .05$), literacy (58.3 vs. 61.9%, $t = 4.16$, $p < .001$), and numeracy (53.8 vs. 57.8%, $t = 4.16$, $p < .001$). They were more likely (14.1% vs. 5.1%) to be in a combined KG classroom ($\chi^2(2) = 52.19$, $p < .001$).

**Teacher/classroom sample.** The majority of schools had two KG teachers, though the range was 1–5. Two teachers (one KG1 and one KG2 teacher) were randomly selected from each school, bringing the total sample in the fall to 444 teachers. By the spring, 107 teachers were no longer teaching KG at the school (which is not surprising given that Ghana’s education system experiences high rates of teacher mobility and turnover; Osei, 2006). In these cases, the new teacher for the same classroom was assessed in the spring. In addition, by the spring, a few schools that had two KG classrooms in the fall had combined them to one classroom due to low enrollment, while a few of the schools that initially had one combined classroom split that class into two levels (KG1 and KG2) when enrollment increased. Finally, in the spring follow-up assessment, four schools dropped out of the study. Thus, the final analytic sample includes the 438 teachers / classrooms from 236 schools ($n$ = 128 private and $n$ = 108 public) that were present in the spring. Sample characteristics are shown in Table 1.

**Measures**

**Child outcomes**. Children's early academic and social-emotional skills were assessed in the fall and spring using the International Development and Early Learning Assessment (IDELA; Pisani et al., 2015; Wolf et al., 2017). Fall scores were included as covariates and spring scores were included as outcome variables. The IDELA is a multi-domain, direct assessment of children's development that was designed for use in global, low-resourced settings. The tool takes approximately 35 minutes to administer and was derived from several commonly used assessments of school readiness, including the Early Development Instrument (EDI; Janus et al., 2007), the Ages and Stage Questionnaire (ASQ; Squires & Bricker, 2009), the Malawi Development (MDAT; Gladstone et al., 2010), and the East Asia Pacific-Early Child Development Scales (EAP-ECDS; Rao et al., 2014). (Further details on the development and implementation of the IDELA can be found in Pisani et al., 2015.)

The factor structure of the IDELA was previously validated in Ethiopia (Wolf et al., 2017). Four domains of the IDELA were used in the present study. Specifically, *Emergent Numeracy* included 39 items across eight subtasks that reflect children's number knowledge, basic addition and subtraction, one-to-one correspondence, shape identification, sorting abilities based on color and shape, size and length differentiation, and completion of a simple puzzle ($\alpha$ = .72 and .70 at fall and spring, respectively). For one-to-one correspondence, for example, children are provided a pile of beans and asked by the assessor to hand him/her a certain number of beans (e.g., 3, 8). *Emergent Literacy* included 38 items across six subtasks that reflect children's print awareness, letter knowledge, phonological awareness, oral comprehension, emergent writing, and expressive vocabulary ($\alpha$ = .74 and .72). For letter knowledge, for example, children are shown a series of letters. The assessor points to one letter at a time and asks the child, "what letter is this?" *Prosocial Skills* (originally called "social-emotional development" in the IDELA) included 14 items across five subtasks that measure children's self-awareness, emotion identification, perspective taking and empathy, friendship, and conflict/problem solving ($\alpha$ = .67 and .70). For perspective taking and empathy, children are shown a drawing of a crying girl and asked to answer questions such as "how do you think this child is feeling right now?" and "what would you do to help her feel better?" *Executive Function* included ten items across two subtasks measuring children's working memory (i.e., forward digit span) and impulse control (i.e., head-toes task; $\alpha$ = .84 and .83). Scores for these domains were calculated as the proportion of items correct, with a possible range from 0 to 1.

The version of the IDELA used in the present study was reviewed by Ghanaian child development experts and piloted with 20 children to ensure cultural applicability. Few minor changes in wording (e.g., simplifying instructions by removing redundant words) were made prior to implementation, but no substantive changes were required. Translations of the IDELA from English into the three local languages (Twi/Fanti, Ewe, and Ga) were conducted iteratively using forward and backward translation conducted by separate individuals, followed by separate conversations with local experts to resolve discrepancies and confirm accuracy. Scoring rules for two of the prosocial items were changed to allow children to receive a "correct" score for more than one response that local experts decided to be relevant within the local context. (For details about IDELA scoring, see Pisani et al., 2015).

A fifth developmental domain – children's approaches to learning – was reported by the assessor. After the assessor completed the IDELA items with each child, s/he filled out seven items about the child's approaches to learning. Each child was rated on a scale of 1 to 4, with 1 = "almost never" and 4 = "almost always." Assessors reported on children's attention (e.g., "Did the child pay attention to the instructions and demonstrations through the assessment?"), confidence, concentration, diligence, pleasure, motivation, and curiosity during the tasks ($\alpha$ = .94 and .92, respectively). Scores for approaches to learning were calculated by taking the average of all items, for a possible range of 1 to 4.

For the present study, scores from the literacy and numeracy domains represent children's early academic skills. Children's scores on the prosocial, executive function, and approaches to learning domains represent social-emotional skills.

**Classroom quality**. The primary predictor for this study – changes in classroom process quality – was measured using the Teacher Instructional Practices and Processes System (TIPPS; Seidman et al., 2013; Wolf, Raza et al., 2018). The TIPPS is an observational tool designed to assess the nature of teacher-child interactions in LMICs. The TIPPS is similar to existing measures of classroom process quality (e.g., the CLASS, ECERS), yet was designed specifically with low-resourced contexts in mind. It was designed to capture theoretical constructs of process quality that promote to child development (e.g., use of structured free play, developmentally-focused instructional practices, classroom management) and are salient in low-income countries (e.g., gender parity). We used the TIPPS-ECE version and made minor adaptations for use in Ghana (e.g., referring to pupils as children, as is common in Ghanaian KG settings). (More information on the tool can be obtained by referring to Seidman et al., 2013; Wolf, Raza et al., 2018).

The TIPPS is comprised of 19 items. We dropped four items due to lack of variability in their scores across classrooms. The four items that were dropped were: Item 1 (“teacher supports children’s development through the use of free playtime;” 100% of classrooms had a rating of 1), Item 2 (“teacher structures learning activities to aide children to learn to work, play and share with others,” 94.0% of classrooms had a rating of 1), Item 15 (“teacher actively employs gender responsive strategies;” 98.4% of the classrooms had a rating of 2), and Item 16 (“teacher actively employs responsive strategies for diverse learners;” 98.7% of classrooms had a rating of 2).

Using exploratory and confirmatory factor analysis (see Wolf, Raza et al., 2018 for details), the remaining 15 items were grouped in to three factors: *Facilitating Deeper Learning* (FDL; 3 items; connecting lesson to teaching objectives, provides specific, high quality feedback, and uses scaffolding; $\alpha$ = .62 and .42 in the fall and spring, respectively); *Supporting Student Expression* (SSE; 4 items; considers student ideas and interests, encourages students to reason and problem solve, connects lesson to students’ daily lives, and models complex language; $\alpha$ = .70 and .64), and *Emotional Support and Behavior Management* (ESBM; 7 items, positive climate, negative climate, sensitivity and responsiveness, tone of voice, positive behavior management, provides consistent routines, student engagement in class activities; $\alpha$ = .70 and .83). The three-factor model had adequate goodness of fit ($\chi^2$ (df) = 154.19 (87), *RMSEA* = .070, *CFI* = .946, *TLI* = .934). See Wolf, Raza et al. (2018) for details on the analysis and concurrent validity of the three factors in this sample. Item scores were averaged within factors to create domain scores that ranged from 1 to 4, with higher scores indicating greater quality.

In addition to the TIPPS, we used a researcher-developed *curriculum checklist* of instructional practices delivered in the classroom during the observation period. The checklist consisted of 13 activities that are described in Ghana’s national ECE curriculum and were explicitly emphasized in the parent study’s teacher training intervention. Items targeted behavior management and instructional practice, and included: “Teacher praises children for positive behavior;” “Teacher threatens children with or used a cane on children at least once (reverse coded);” “Teacher explicitly reminds children of the class rules;” “Teacher uses a signal to gain children’s attention (e.g., drum beat, song, bell);” “Children are seated in a way that children can see each other’s faces (e.g., in a circle, or tables together in groups);” “Teacher uses one or multiple songs to facilitate learning at some point in the lesson;” and “There is an activity that facilitated the lesson objectives that involved manipulation of materials.” Each item was coded as either present in the video (a score of 1) or absent in the video (a score of 0), and overall scores represented item sums with a potential range of 0 to 13.

For the present study, FDL and the curriculum checklist were used to represent instructional components of classroom quality, whereas ESBM and SSE were used to represent social-emotional aspects of classroom quality. Classroom scores on each of the four quality domains in the fall were included as covariates. The primary predictor variables in the present study were classroom change scores, which were calculated by subtracting the fall score from the spring score for each quality domain. Positive numbers represent improvements in classroom quality, whereas negative scores represent reductions in classroom quality over the eight-month period. Fall level and change scores were significantly and negatively associated within quality domain, with correlations ranging from $r$ = -.71 for FDL to $r$ = -.58 for the curriculum checklist, indicating that classrooms that started at lower levels improved more over the year. Correlations across domains for level and change scores were substantially lower (all < .4). See Appendix Table 1 for correlations between study variables.

**Covariates**. In addition to controlling for fall IDELA and classroom quality scores, several additional covariates were used in the analyses to account for possible confounding characteristics. Children's age and gender were reported by primary caregivers at the beginning of the school year and confirmed by data collectors during the child assessment. Primary caregivers also reported on their own age (a proxy for teenage parenthood), their marital status (married vs. not), and a set of ten household variables (e.g., number of household members, highest grade completed by female head or spouse, employment of male head of house, materials used for construction of household's roof, source of drinking water, possession of materials such as working radio). These ten variables were combined to construct the Simple Poverty Scorecard for Ghana (Schreiner & Woller, 2010), a validated metric of household poverty levels for Ghana that ranges from 0 to 100 (with lower scores indicating less wealth / higher poverty). Three additional school-related variables were included: an indicator for whether the school was private (vs. public), an indicator for whether the classroom experienced a change in the teacher from fall to spring, and a set of indicators for the school's district.

## Analytic Plan

**Missing data**. Although no children in our analytic sample were missing outcome data, a non-negligible proportion were missing other predictor variables and covariates (see Table 1). In particular, a substantial percentage of cases (28.5%) were missing on classroom quality data in the fall. Most (75.2%) of the missing fall classroom data were attributable to the theft of three tablets and two netbook computers containing 94 classroom videos, and therefore can be considered missing completely at random. Approximately 23 percent of household covariates (caregiver age, caregiver marital status, and household poverty) were also missing due to non-response of caregivers to a phone-based survey.

We used multiple imputation (with Stata's "ice" command) to handle missing data on covariates and independent (but not dependent) variables. We follow recommended best practice by Johnson and Young (2011) by using the full set of covariates, as well as the dependent variables, during imputation modeling. All estimates (calculated using Stata's "mi estimate" command) were then derived using data from 20 multiply imputed datasets.

**Aim 1: Main effects of changes in classroom quality**. To understand whether changes in classroom quality across the school year predicted gains in child outcomes (Aim 1), we used a multi-level regression equation with children nested in classrooms nested in schools, represented in Equation 1:

*Level 1:* Child

$$Y_{ijk} = \pi_{0jk} + \sum_{n} \pi_n X_{ijk} + e_{ijk}$$

*Level 2:* Teacher/Classroom

$$\pi_{0jk} = \beta_{0k} + \sum_{m} \beta_m T_{jk} + \sum_{p} \beta_p Change_{jk} + r_{jk}$$

*Level 3:* School

$$\beta_{0k} = \gamma_0 + \sum_{q} \gamma_q S_k + u_k$$

where:

$Y_{ijk}$ = the value of the given spring outcome score (literacy, numeracy, prosocial, executive function, approaches to learning) for child *i* in classroom *j* in school *k;*

$X_{ijk}$ = the value of baseline covariate *n* for child *i* in classroom *j* in school *k* (including child fall outcome score, child gender, child age, caregiver age, caregiver marital status, household poverty level);

$T_{jk}$ = the value of baseline covariate *m* for classroom *j* in school *k* (including fall levels of FDL, ESBM, SSE, and the curriculum checklist, as well as whether the classroom experienced a change in the teacher over the school year);

$Change_{jk}$ = the change score for classroom quality variable $p$ within classroom $j$ and school $k$ (including changes in FDL, ESBM, SSE, and the curriculum checklist across the school year);

$S_k$ = the value of fall covariate $q$ for school $k$ (including whether the school is private vs. public, as well as the district in which the school is located)

$\gamma_0$ = the overall intercept;

$e_{ijk}$ = a random error that varies across individuals;

$r_{jk}$ = a random error that varies across classrooms/teachers; and

$u_k$ = a random error that varies across schools.

To address Aim 1, we focus on the coefficients for each classroom quality change variable ($\beta_p$).

**Aim 2: Interactions with fall quality and child skill levels**. To understand whether the associations between changes in classroom quality and child outcomes differed based on initial (fall) levels of classroom quality (Aim 2a) or based on initial (fall) levels of child skill (Aim 2b), we added several different interaction terms to the above set of equations. First, for Aim 2a, we fit a set of regression models that included interactions between *Level* and *Change* (at the classroom level) for each classroom quality domain. Next, for Aim 2b, we fit a set of regression models that included cross-level interactions between *Change* (at the classroom level) and children's fall outcome score (at the individual child level). For interaction models, all primary predictor and moderator variables were grand mean centered.

**Sensitivity analyses**. We conducted several sensitivity analyses to test the robustness of our primary models against alternative model specifications. First, we run our analyses using fixed effects for schools and clustered standard errors for classrooms rather than a multi-level model. Second, we replicated our models using listwise deletion instead of multiple imputation. Third, we replicated our models within the parent study's control and treatment groups separately. Descriptive analyses suggest that classrooms in the treatment groups showed larger and more positive average changes in classroom quality over the course of the year (e.g., *M* change for FDL = 0.29, for ESBM = 0.29, for SSE = 0.28, and for curriculum checklist = 1.45) compared to those in the control group (e.g., *M* change for FDL = 0.12, for ESBM = 0.20, for SSE = -0.05, and for curriculum checklist = -0.03). Although our primary interest was in understanding the role that change in classroom quality – induced by any means – has in supporting children's outcomes, we focus on these separate groups to ensure that results were not being driven by participation in the study's intervention. Fourth, we replicated the primary models excluding classrooms that experienced teacher turnover across the year. Although the main models included a covariate for teacher turnover, these sensitivity analyses were intended to ensure that the results were driven by children's exposure to changes in classroom quality, rather than changes in specific teachers. Finally, we re-ran all models including each domain of classroom quality individually. The literature is mixed with regard to whether classroom quality domains should be analyzed collectively or independently (e.g., Burchinal et al., 2010; Hamre & Pianta, 2005; Mashburn et al., 2008; Weiland et al., 2013). Our primary models include all domains together in order to isolate the predictive power of the unique variance associated with each domain. This is appropriate given that the correlations between domains observed in this study ($r$'s < .40) are comparatively weak relative to those found in prior work (e.g., Weiland et al., 2013, which showed $r$'s of .70 and above). At the same time, we also present results in which domains are included individually to complement prior research using this approach, as well as to assess the possibility of suppression effects due to collinearity.

# Results

## Descriptive Statistics

Results of descriptive analyses are shown in Table 1. On average, classrooms showed improvements – rather than declines – in all quality domains over the course of the school year. Using a 1 to 4 scale, classrooms showed an average improvement of 0.21 points in FDL, 0.26 points in ESBM, and 0.17 points in SSE. Classrooms also demonstrated an average of 1.00 more nationally-supported curricular activity (out of 13) according to the curriculum checklist. Despite these improvements, classrooms' average spring SSE scores (1.72 on a scale of 1 to 4) remained particularly low compared to classrooms' spring FDL (2.36) and ESBM (3.05) scores. Overall number of curricular activities at the end of the year was also low, at an average of 4.07 activities completed out of 13. Children's developmental scores showed average improvements across the year in all domains, with effect sizes for these gains ranging from 0.17 *SDs* for approaches to learning to 0.74 *SDs* for literacy.

## Aim 1: Main Effects of Classroom Quality Changes

Results of analyses examining the association between changes in classroom quality over the course of the school year and gains in children's developmental outcomes are shown in Table 2. In all tables and in the text below, we present unstandardized coefficients ($b$) representing the association in the measures' original units, as well as standardized coefficients ($\beta$) representing the association in standard deviation units. Positive changes in FDL from the beginning to the end of the year were significantly and positively associated with gains in children's literacy skills, $b = 0.006$, S.E. = 0.003, $p < .05$, $\beta = 0.030$, and their prosocial skills, $b = 0.010$, S.E. = 0.004, $p < .01$, $\beta = 0.052$. They were also positively associated with children's executive function skills at the trend level, $b = 0.007$, S.E. = 0.003, $p < .10$, $\beta = 0.035$. Changes in FDL were not significantly associated with differences in numeracy or approaches to learning skills.

Positive changes in ESBM were significantly and positively associated with gains in numeracy skills, $b = 0.012$, S.E. = 0.005, $p < .05$, $\beta = 0.034$. Positive changes in ESBM were negatively associated with children's prosocial skills, $b = -0.013$, S.E. = 0.006, $p < .05$, $\beta = -0.038$, and their approaches to learning, $b = -0.041$, S.E. = 0.019, $p < .05$, $\beta = -0.036$. Changes in ESBM were not significantly associated with differences in literacy or executive function skills.

Next, we found that positive changes in SSE were significantly and negatively associated with children's literacy skills, $b = -0.007$, S.E. = 0.003, $p < .05$, $\beta = -0.028$, and their approaches to learning skills, $b = -0.022$, S.E. = 0.011, $p < .05$, $\beta = -0.029$. They were also negatively associated with children's prosocial skills at the trend level, $b = -0.006$, S.E. = 0.004, $p < .10$, $\beta = -0.026$. Changes in SSE were not significantly associated with differences in numeracy or executive function skills.

Finally, positive changes in the curriculum checklist were significantly and positively associated with gains in children's literacy skills, $b = 0.006$, S.E. = 0.001, $p < .001$, $\beta = 0.061$, numeracy skills, $b = 0.002$, S.E. = 0.001, $p < .05$, $\beta = 0.024$, prosocial skills, $b = 0.005$, S.E. = 0.001, $p < .001$, $\beta = 0.057$, executive function skills, $b = 0.003$, S.E. = 0.001, $p < .05$, $\beta = 0.037$, and their approaches to learning, $b = 0.020$, S.E. = 0.004, $p < .001$, $\beta = 0.068$.

## Aim 2: Differences by Fall Level Scores

**Fall classroom quality.** Results revealed some evidence for differences in the associations between changes in classroom quality and child outcomes based on fall levels of classroom quality (Aim 2a). Of the 24 interactions tested, 11 were significant at $p < .10$. Overall, results suggested that children benefited more from improvements in FDL and ESBM in classrooms with higher starting levels of these quality characteristics, although the significance of these differences varied across outcomes. In particular, the associations between changes in FDL and child outcomes were significantly more positive in classrooms with higher levels of fall FDL for children's numeracy skills, $b = 0.002$, S.E. = 0.001, $p < .05$, $\beta = 0.007$, prosocial skills, $b = 0.004$, S.E. = 0.001, $p < .001$, $\beta = 0.015$, and approaches to learning, $b = 0.007$, S.E. = 0.004, $p < .05$, $\beta = 0.009$, but differences were not observed for literacy and executive function. For ESBM, the relation between changes in quality and children's executive function was significantly more positive for classrooms with higher starting levels of ESBM, $b = 0.012$, S.E. = 0.003, $p < .001$, $\beta = 0.018$. No additional differences based on initial ESBM levels were observed for any other child outcome.

Results also suggested that children benefited more from improvements in SSE and the curriculum checklist in classrooms with lower starting levels of quality in these areas. Once again, the patterns of differences varied across child outcomes. Specifically, the associations between changes in SSE and child outcomes were significantly more positive in classrooms with lower levels of fall SSE for approaches to learning, $b = -0.012$, S.E. = 0.006, $p = .05$, $\beta = -0.012$, and at the trend level for numeracy skills, $b = -0.003$, S.E. = 0.001, $p < .10$, $\beta = -0.009$. No differences in associations were observed based on fall levels of SSE for literacy, prosocial, or executive function skills. Finally, the associations between changes in the curriculum checklist and children's outcomes were significantly more positive in classrooms with lower levels of curriculum use for children's literacy skills, $b = -0.0004$, S.E. = 0.0002, $p < .05$, $\beta = -0.007$, and executive function, $b = -0.001$, S.E. = 0.000, $p < .05$, $\beta = -0.011$, and at the trend level for prosocial skills, $b = -0.0004$, S.E. = 0.0002, $p < .10$, $\beta = -0.008$. No differences in associations were observed based on initial curriculum use for children's numeracy or approaches to learning skills.

**Fall child outcomes.** Results revealed stronger evidence for differences in associations between changes in classroom quality and child outcomes based on children's fall developmental skill levels (Aim 2b). Of the 24 interactions tested, 22 were significant at $p < .10$. Overall, results suggested that children benefited more from improvements in FDL, SSE, and the curricular checklist when they entered school with higher levels of skills, although the consistency of these results varied. In particular, the associations between changes in FDL quality and gains in children's outcomes were significantly more positive for children with higher levels of baseline literacy skills, $b = 0.012$, S.E. = 0.003, $p < .001$, $\beta = 0.012$, prosocial skills, $b = 0.015$, S.E. = 0.004, $p < .001$, $\beta = 0.015$, and approaches to learning, $b = 0.022$, S.E. = 0.003, $p < .001$, $\beta = 0.025$, and at the trend level for numeracy skills, $b = 0.006$, S.E. = 0.003, $p < .10$, $\beta = 0.006$. The association between changes in FDL quality and gains in children's executive function skills, however, was more positive for children with lower fall skills, $b = -0.019$, S.E. = 0.004, $p < .001$, $\beta = -0.020$. For SSE, the associations between changes in classroom quality and gains in children's outcomes were significantly more positive for children with higher baseline levels of literacy skills, $b = 0.019$, S.E. = 0.004, $p < .001$, $\beta = 0.016$, numeracy skills, $b = 0.034$, S.E. = 0.004, $p < .001$, $\beta = 0.028$, prosocial skills, $b = 0.022$, S.E. = 0.004, $p < .001$, $\beta = 0.018$, and executive function, $b = 0.022$, S.E. = 0.004, $p < .001$, $\beta = 0.020$. The association between changes in SSE quality and gains in children's approaches to learning skills, on the other hand, was more positive for children with lower fall skills, $b = -0.018$, S.E. = 0.004, $p < .001$, $\beta = -0.017$. Finally, for the curriculum checklist, the associations between changes in quality and gains in child outcomes were significantly more positive for children with higher levels of fall numeracy skills, $b = 0.003$, S.E. = 0.002, $p < .05$, $\beta = 0.007$, prosocial skills, $b = 0.011$, S.E. = 0.002, $p < .001$, $\beta = 0.024$, and approaches to learning skills, $b = 0.004$, S.E. = 0.001, $p < .01$, $\beta =$

0.011. Like FDL, however, the association between gains in curriculum checklist scores and children's executive function skills was more positive for children with lower fall skills, $b$ = -0.006, S.E. = 0.002, $p < .001$, $\beta = -0.014$. No differences in benefits attributable to curriculum checklist gains were found for children's literacy skills.

Results for ESBM showed different patterns, with children appearing to benefit more from improvements in ESBM quality when they entered school with lower skill levels. Specifically, associations between changes in ESBM and gains in children's outcomes were more positive for children with lower levels of baseline numeracy skills, $b$ = -0.017, S.E. = 0.006, $p < .01$, $\beta = -0.009$, prosocial skills, $b$ = -0.037, S.E. = 0.007, $p < .001$, $\beta = -0.021$, executive function skills, $b$ = -0.032, S.E. = 0.006, $p < .001$, $\beta = -0.019$, and approaches to learning skills, $b$ = -0.018, S.E. = 0.005, $p < .01$, $\beta = -0.012$.

**Sensitivity Analyses**

Results of sensitivity analyses are shown in Appendix Table 2. The results of sensitivity analyses (1) using a fixed effects specification, (2) using listwise deletion in place of multiple imputation, (3) replicating the primary models within the control and treatment groups, and (4) replicating the primary models within the subsample of children whose teachers were the same across fall and spring were largely consistent in terms of magnitude and direction of coefficients compared with those from the primary models. Some exceptions, however, were found. Changes in SSE, for example, were found to negatively predict children's executive function in the control group, $b$ = -0.011, S.E. = 0.006, $p < .10$, $\beta = -0.050$, and to positively predict these skills in the treatment group, $b$ = 0.014, S.E. = 0.005, $p < .01$, $\beta = 0.060$. Furthermore, given reductions in statistical power in many of the sensitivity analyses, fewer of these associations were statistically significant. Findings for the curriculum checklist were most robust across the sensitivity analyses, showing positive and statistically significant associations with child outcomes for most results.

For our final set of sensitivity analyses examining each domain of classroom quality in a separate (rather than combined) model, results were largely consistent for FDL and the curriculum checklist. Notably, however, these domain-separated models did not identify any statistically significant, negative associations between changes in ESBM or SSE quality and children's outcomes, as observed in the primary models. Instead, changes in ESBM were shown to predict positive, trend-level gains in children's literacy skills, $b = 0.010$, S.E. = 0.005, $p < .10$, $\beta = 0.026$, and executive function skills, $b = 0.011$, S.E. = 0.006, $p < .10$, $\beta = 0.031$, whereas changes in SSE were shown to predict positive and statistically significant gains in children's executive function, $b = 0.010$, S.E. = 0.003, $p < .001$, $\beta = 0.045$. These findings suggest that negative associations between changes in social-emotional dimensions of classroom quality and child outcomes only exist when accounting for both levels and changes in instruction-oriented dimensions of quality.

## Discussion

The aim of this study was to examine changes in different dimensions of preschool classroom process quality from the beginning to the end of one school year as predictors of Ghanaian children's gains in early academic and social-emotional skills. Our results suggest that improvements in instruction-oriented dimensions of classroom quality predicted positive gains not only in children's early academic skills, but also in their social-emotional outcomes. Contrary to our hypothesis, when accounting for levels and changes in instructional quality, improvements in social-emotional dimensions of quality were negatively associated with children's prosocial and approaches to learning skills, not related to children's executive function skills, and inconsistently related to their early academic skills. Notably, the magnitudes of all associations were small (with standardized coefficients all falling below 0.07), which is consistent with prior research on classroom quality that has attempted to account for selection bias (see Burchinal, 2017).

In the present study, improvements in teachers' use of pedagogical practices set forth by the national KG curriculum (e.g., use of manipulatives, attention signals, songs) and their facilitation of deeper learning were positively associated with gains in not just students' learning outcomes, but also their social-emotional skills. The size of these coefficients was small ($\beta$ = 0.02 – 0.07), with gains in the classroom checklist, for example, accounting for anywhere from four percent of children's gains in numeracy skills to 38 percent of children's gains in approaches to learning across the school year. While the results indicate that measured dimensions of classroom quality do reliably predict child outcomes over the course of one school year, they also point to the potentially large role of unmeasured representations of classroom quality and contexts outside of the classroom in driving children's learning outcomes (Mashburn, 2017; Wagner, 2018).

On the one hand, these results support and extend a body of mostly cross-sectional work from high-income countries highlighting positive associations between instructional support and children's learning outcomes, and especially their language and literacy skills (Burchinal et al., 2008; Howes et al., 2008; Mashburn et al., 2008). On the other hand, the positive associations between these instructional quality characteristics and child social-emotional outcomes is not necessarily consistent with prior work from the U.S., which has typically shown social-emotional outcomes to be significantly predicted only by emotional climate (Burchinal et al., 2010) or by nothing at all (Keys et al., 2013). These findings suggest there may be something unique either about the educational context of Ghana or about the measures used to operationalize instructional quality that supports cross-domain relations. In particular, the curriculum checklist measure included both instructional approaches *and* practices related to classroom management, which may have had implications for children's social and emotional wellbeing.

Our results also suggest that when accounting for instructional quality, improvements in observed emotional support, behavior management, and support of student expression had, on average, negative associations with several dimensions of children's social-emotional wellbeing. Once again, the magnitudes of these associations were quite small ($\beta = 0.03 – 0.04$). Although the links in the literature between these relational dimensions of classroom quality and children's outcomes are less well established than the links for instructional quality, these results are nevertheless surprising. Indeed, cross-sectional evidence from high-income countries has shown that teachers who are warm and responsive to children's needs, and who use positive behavioral management strategies tend to have children with fewer behavioral challenges (Mashburn et al., 2008). Also surprising was the fact that despite negative associations with social-emotional outcomes, we found positive associations between gains in emotional support and behavior management techniques and children's early numeracy skills. Experimental evidence from the U.S.-based Chicago School Readiness Project (CSRP) found positive impacts of a teacher behavior management intervention on children's early numeracy knowledge (Raver et al., 2011). As such, this finding may indicate additional potential for cross-domain relations between classroom practices that support (potentially unmeasured) social-emotional skills such as self-regulation and children's academic gains (McClelland et al., 2007) in the Ghanaian educational context.

There are several reasons why we may be observing these patterns of results related to social-emotional climate. First, it is important to note that we did observe a slightly different pattern of results in our sensitivity analyses examining improvements in social-emotional quality dimensions individually. In these analyses, classroom gains in support of student expression and in emotional support and behavior management predicted statistically significant, positive gains in children's executive function, and did not show negative associations with social-emotional outcomes. As such, our results do not necessarily imply that teachers' emotional interactions with students are unimportant or harmful. Rather, they indicate that the effort or strategies required to improve classrooms' social-emotional quality *above and beyond* other forms of student-teacher interactions may have unintended negative consequences for children's outcomes (e.g., by increasing teachers' stress or taking time away from other social or emotional learning opportunities). Alternatively, it is also possible that our findings reflect reverse causality. For example, rather than changes in classroom quality driving changes in student behavior, it is possible that teachers in this sample responded to behavior problems in their classrooms by using more supportive teaching practices.

Furthermore, the mixed findings with regard to social-emotional elements of classroom quality may reflect a lack of appropriate measurement of culturally-salient social-emotional processes. African scholars have noted the importance of the social ontogenetic paradigm (Nsamenang, 2005) in understanding early child development in Africa. This is premised on interdependent relationships, with socialization being organized to teach social competence and shared responsibly within the family and community (rather than for individualization or academic pursuits; Nsamenang & Lamb, 1994). No research to date has considered if and how the social ontogenetic paradigm is integrated into school and classroom environments, and none of the measures used in this study capture this dimension of classrooms. Understanding the ways that teachers promote children's development in SSA classrooms and in other collectivist cultures is a needed area for future research.

In addition to exploring the overall associations between gains in classroom quality and child outcomes over one school year, we also examined the degree to which these processes varied based on initial levels of quality and/or child skill at the beginning of the school year. Although effect sizes were once again small, nearly 70 percent of interactions tested were statistically significant, indicating relatively consistent evidence for differences. Overall, these results suggested that the benefits of improvements in classroom quality were more positive in classrooms where teachers showed higher initial levels of facilitating deeper understanding and emotional support and behavior management, and in classrooms where teachers initially struggled with supporting student expression and implementing instructional strategies set forth by the national curriculum. We also found that improvements in facilitating deeper learning, supporting student expression, and implementing curricular strategies were more positively associated with children's outcomes when students entered KG with higher levels of skills. Benefits associated with emotional support and behavior management, on the other hand, were strongest for children entering with lower skill levels. Importantly, these results were relatively inconsistent across child outcomes, suggesting substantial heterogeneity in these processes.

Together, these findings reinforce the complexity of classroom processes and the ways in which they might differentially benefit children. For example, although teachers' increased use of instructional practices such as scaffolding to facilitate students' deeper learning predicts learning for all children in the sample on average, its benefits appear to accrue most in classrooms and for students with relatively high levels of baseline proficiency. This finding supports a hypothesis of "accumulated advantage," whereby children with greater skill levels or environmental supports are better prepared to take advantage of future learning experiences (Stanovich, 1986). Similar patterns of accumulated advantage have been observed in other studies of early childhood programs in the United States and Europe (Lehrl, Kluczniok, & Rossbach, 2016). These results also complement research showing added benefits of cross-sectional increments in instructional quality for high-quality classrooms only (Burchinal et al., 2010; Leyva et al., 2015; Hatfield et al., 2016). These findings suggest that interventions in this setting to support more advanced instructional practices should be prioritized in high-performing classrooms, and that instructional interventions implemented in lower-performing settings may require additional, more basic supports to optimize effectiveness.

Our findings in other areas of classroom quality, however, did not necessarily support a conclusion of accumulated advantage. For example, children whose teachers increased their support of student expression tended to show either decreases or no changes in their outcomes over the school year. Importantly, however, changes in support of student expression appeared to be more beneficial (or, in the case of some outcomes, less negative) for children who came from classrooms where teachers initially struggled relatively with these skills. This finding is consistent with a compensatory hypothesis, which posits that the most disadvantaged individuals have the most “room to grow,” and therefore stand to benefit most from environmental supports. This hypothesis has also shown support in the early childhood literature (McCartney, Dearing, Taylor, & Bub, 2007; Weiland & Yoshikawa, 2013) and is the basis for many publicly-funded programs that aim to “level the playing field” for low-income children (e.g., Head Start). Our results suggest that interventions to help Ghanaian teachers to consider student ideas and interests and to connect lessons to students’ daily lives should be prioritized for teachers struggling with these skills. At the same time, as noted above, our findings of overall null or negative associations between improvements in these forms of quality and children's outcomes imply that alternative student support strategies that were not assessed in this study may lead to more optimal outcomes within this particular cultural context.

**Limitations**

Findings must be interpreted within the context of the study's limitations. First, although our use of classroom quality change scores and inclusion of baseline child skill levels as covariates strengthen the internal validity of our estimates (by accounting for time-invariant differences across classrooms/children), the results of these analyses cannot be considered causal. Our failure to account for all potential omitted variables may have led us to misestimate the association between classroom quality and children's developmental outcomes. Future research leveraging experimental and quasi-experimental methods (e.g., instrumental variables) is needed to confirm the causality of these relations.

Next, although our measures of both classroom quality and child outcomes were carefully tested for their appropriateness within the Ghanaian cultural context, as noted above, it is likely that they do not cover all pedagogical and developmental processes relevant within this setting. Research has shown that many SSA cultures place a strong emphasis on children's development of social responsibility and other relational skills not fully covered by the IDELA (e.g., Serpell, 2011). Although less is known about alternative conceptualizations of classroom quality in this region, it is likely that the TIPPS does not capture all potentially relevant dimensions of quality. Furthermore, the low internal consistency for our measure of FDL in the spring suggests that this domain may not have been measured reliably at this time point. Finally, the fact that the TIPPS was collected only once in the beginning and the end of the school year does not allow us to untangle specific patterns of change over time, nor does it permit analyses to understand differences in quality within the school day (e.g., across different types of activities). Indeed, a larger set of observations for each time point, and periods of observations throughout the school year, would strengthen the results. Future research should endeavor to deepen the frequency and cultural specificity of measures of quality in diverse parts of the world, while also attending closely to psychometric performance.

Finally, despite the fact that this study was conducted in a setting that is under-represented in academic research on classroom quality, the findings presented here cannot be generalized outside of the peri-urban Greater Accra Region. Research in SSA has noted the vast differences in the risk and protective factors experienced by children living in urban and rural settings (Zhang, 2006), and in Ghana specifically (Cooke et al., 2016). Differences in cultural and linguistic processes also limit the degree to which these findings might generalize to other countries within the region. As such, additional research is needed to understand the external validity of these associations in diverse LMICs and, in particular, across SSA.

## Conclusions and Implications

The results of the present study suggest that improvements in Ghanaian preschool classrooms' instructional quality are related to small, positive gains in children's early academic and social-emotional outcomes from the beginning to the end of one school year, and that these improvements may be particularly beneficial for children and classrooms with higher baseline proficiency levels. Evidence for the benefits of changes in classroom social-emotional practices, however, is mixed. In the future, additional research is needed to understand whether alternative approaches to behavior management and student-teacher interaction may be more beneficial in supporting children in diverse cultural contexts. Furthermore, there may be a need for observational tools that assess how teachers promote a social ontogenetic paradigm (Nsamenang,

2005) within classrooms in SSA.

This work builds on and extends prior work from high-income countries showing cross-sectional associations between classroom processes and children's learning. In particular, these findings reinforce the importance of policy efforts aimed not only at expanding access to preschool programming in LMICs, but also at ensuring that these programs adhere to high standards for instructional quality. At the same time, the small magnitudes of the associations observed in this study suggest that additional or alternative strategies may be necessary for generating transformational changes in children's learning. For example, continued implementation and expansion of pre- and in-service teacher training programs that support teachers' pedagogical practice and rigorous application of evidence-based, culturally-relevant curricula may be necessary to generate the types of changes in classroom quality that are likely to generate noticeable benefits in children's learning outcomes. Alternatively, these small effect sizes may reflect challenges associated with precisely measuring relevant classroom processes in LMICs. Future research is needed to improve the measurement of classroom quality, as well as to understand the degree to which these findings generalize to diverse LMIC and high-income settings where longitudinal research on classroom quality is lacking. It is with this evidence base that ECE expansion across LMICs can include targeted support and training for teachers to ensure high quality, and ultimately improve developmental opportunities for all children.

**References**

Aboud, F. E. (2006). Evaluation of an early childhood preschool program in rural Bangladesh. *Early Childhood Research Quarterly*, *21*(1), 46-60, http://dx.doi.org/10.1016/j.ecresq.2006.01.008

Akyeampong, K. (2017). Teacher educators' practice and vision of good teaching in teacher education reform context in Ghana. *Educational Researcher*, *46*(4), 194-203.

Angrist, J. D., Imbens, G.W., & Ruin, D.R. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association, 91(434),* 444-455. http://dx.doi.org/10.2307/2291629

Araujo, M. C., Carneiro, P. M., Cruz-Aguayo, Y., & Schady, N. (2016). *Teacher Quality and Learning Outcomes in Kindergarten*. IDB Working Paper Series, # IDB-WP-665.

Auger, A., Farkas, G., Burchinal, M. R., Duncan, G. J., & Vandell, D. L. (2014). Preschool center care quality effects on academic achievement: An instrumental variables analysis. *Developmental Psychology*, *50*(12), 2559. http://dx.doi.org/10.1037/a0037995

Barnett, W. S., Friedman-Krauss, A. H., Weisenfeld, G. G., Horowitz, M., Kasmin, R., & Squires, J. H. (2017). *The State of Preschool 2016: State Preschool Yearbook*. New Brunswick, NJ: National Institute for Early Education Research.

Bidwell, K., Perry, K., & Watine, L. (2014). *Exploring Early Education Programs in Peri-urban Settings in Africa: Accra, Ghana*. Innovations for Poverty Action: New Haven, CT.

Black, M. M., Walker, S. P., Fernald, L. C., Andersen, C. T., DiGirolamo, A. M., Lu, C., ... & Devercelli, A. (2017). Early childhood development coming of age: Science through the life course. *The Lancet*, *389*, 77-90, http://dx.doi.org/10.1016/S0140-6736(16)31389-7

Bornstein, M. H., Putnick, D. L., Bradley, R. H., Lansford, J. E., & Deater-Deckard, K. (2015). Pathways among caregiver education, household resources, and infant growth in 39 low- and middle-income countries. *Infancy*, *20*, 353-376.

Brinkman, S. A., Hasan, A., Jung, H., Kinnell, A., Nakajima, N., & Pradhan, M. (2017). The role of preschool quality in promoting child development: Evidence from rural Indonesia. *European Early Childhood Education Research Journal*, 1-23, http://dx.doi.org/10.1080/1350293X.2017.1331062

Britto, P. R., Yoshikawa, H., & Boller, K. (2011). Quality of early childhood development programs in global contexts: Rationale for investment, conceptual framework and implications for equity. *Social Policy Report*, *25*(2).

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In R. M. Lerner (Ed.) *Handbook of child development: Vol. 1. Theoretical models of human development* (6th ed., pp. 793 – 828). Hoboken, NJ: Wiley.

Burchinal, M., Howes, C., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Predicting child outcomes at the end of kindergarten from the quality of pre-kindergarten teacher–child interactions and instruction. *Applied Development Science*, *12*(3), 140-153, http://dx.doi.org/10.1080/10888690802199418

Burchinal, M., Kainz, K., & Cai, Y. (2011). How well do our measures of quality predict child outcomes? A meta-analysis and coordinated analysis of data from large-scale studies of early childhood settings. *Quality Measurement in Early Childhood Settings*, 11-31.

Burchinal, M., Xue, Y., Auger, A., Tien, H. C., Mashburn, A., Peisner-Feinberg, E., ... & Tarullo, L. (2016). III. Testing for quality thresholds and features in early care and education. *Monographs of the Society for Research in Child Development*, *81*(2), 46-63. http://dx.doi.org/10.1111/mono.12238

Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in prekindergarten programs. *Early Childhood Research Quarterly*, *25*(2), 166-176, http://dx.doi.org/10.1016/j.ecresq.2009.10.004

Cooke, E., Hague, S., & McKay, A. (2016). The Ghana poverty and inequality report: Using the 6th Ghana Living Standards Survey. New York: UNICEF.

Curby, T. W., LoCasale-Crouch, J., Konold, T. R., Pianta, R. C., Howes, C., Burchinal, M., ... & Barbarin, O. (2009). The relations of observed pre-K classroom quality profiles to children's achievement and social competence. *Early Education and Development*, *20*(2), 346-372, http://dx.doi.org/10.1080/10409280802581284

Ghana Ministry of Education (2016). *Education Sector Performance Report, 2016.* Ghana.

Gladstone, M., Lancaster, G., Umar, E., Nyirenda, M., Kayira, E., van den Broek, N., & Smyth, R. (2010). The Malawi Developmental Assessment Tool (MDAT): The creation, validation, and reliability of a tool to assess child development in rural African settings. *PLoS Medicine*, *7*, e1000273, http://dx.doi.org/10.1371/journal.pmed.1000273

Gonzalez, K. E., McCoy, D. C., Sabol., T. J. (2017). Moving beyond associations: Identifying the causal impact of observed early childhood classroom quality on child outcomes. *Under Review.*

Gordon, R. A., Hofer, K. G., Aloe, A. M., Wilson, S. J., Peng, F., Gaur, D., & Lambouths III, D. (2017). Preschool quality and children's vocabulary: A meta-analytic study of heterogeneity and moderation of effects. Chicago, IL.

Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development*, *76*(5), 949-967, http://dx.doi.org/10.1111/j.1467-8624.2005.00889.x

Hamre, B., Hatfield, B., Pianta, R., & Jamil, F. (2014). Evidence for general and domain-specific elements of teacher–child interactions: Associations with preschool children's development. *Child Development*, *85*, 1257-1274. http://dx.doi.org/10.1111/cdev.12184

Hatfield, B. E., Burchinal, M. R., Pianta, R. C., & Sideris, J. (2016). Thresholds in the association between quality of teacher–child interactions and preschool children's school readiness skills. *Early Childhood Research Quarterly*, *36*, 561-571, http://dx.doi.org/10.1016/j.ecresq.2015.09.005

Herrera, M. O., Mathiesen, M., Merino, J. M., & Recart, I. (2005). Learning contexts for young children in Chile: Process quality assessment in preschool centres. *International Journal of Early Years Education*, *13*, 13-27. http://dx.doi.org/10.1080/09669760500048253

Howes, C., James, J., & Ritchie, S. (2003). Pathways to effective teaching. *Early Childhood Research Quarterly*, *18*, 104–120, http://dx.doi.org/10.1016/S0885-2006(03)00008-5

Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Ready to learn? Children's pre-academic achievement in pre-kindergarten programs. *Early Childhood Research Quarterly*, *23*(1), 27-50, http://dx.doi.org/10.1016/j.ecresq.2007.05.002

Janus, M., Brinkman, S., Duku, E., Hertzman, C., Santos, R., Sayers, M., ... & Walsh, C. (2007). *The Early Development Instrument: A population-based measure for communities*. Offord Centre for Child Studies, McMaster University.

Johnson, D. R., & Young, R. (2011). Toward best practices in analyzing datasets with missing data: Comparisons and recommendations. *Journal of Marriage and Family*, *73*, 926-945, http://dx.doi.org/10.1111/j.1741-3737.2011.00861.x

Keys, T. D., Farkas, G., Burchinal, M. R., Duncan, G. J., Vandell, D. L., Li, W., ... & Howes, C. (2013). Preschool center quality and school readiness: Quality effects and variation by demographic and child characteristics. *Child Development*, *84*(4), 1171-1190. http://dx.doi.org/10.1111/cdev.12048

Kuger, S., Kluczniok, K., Kaplan, D., & Rossbach, H. G. (2016). Stability and patterns of classroom quality in German early childhood education and care. *School Effectiveness & School Improvement*, *27*, 418-440, http://dx.doi.org/10.1080/09243453.2015.1112815

Lehrl, S., Kluczniok, K., & Rossbach, H. G. (2016). Longer-term associations of preschool education: The predictive role of preschool quality for the development of mathematical skills through elementary school. *Early Childhood Research Quarterly*, *36*, 475-488, http://dx.doi.org/10.1016/j.ecresq.2016.01.013

Leyva, D., Weiland, C., Barata, M., Yoshikawa, H., Snow, C., Treviño, E., & Rolla, A. (2015). Teacher–child interactions in Chile and their associations with prekindergarten outcomes. *Child Development*, *86*(3), 781-799, http://dx.doi.org/10.1111/cdev.12342

Mashburn, A. J. (2017). Evaluating the validity of classroom observations in the Head Start designation renewal system. *Educational Psychologist*, *52*(1), 38-49, https://doi.org/10.1080/00461520.2016.1207539

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., ... & Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, *79*, 732-749, http://dx.doi.org/10.1111/j.1467-8624.2008.01154.x

McCartney, K., Dearing, E., Taylor, B. A., & Bub, K. L. (2007). Quality child care supports the achievement of low-income children: Direct and indirect pathways through caregiving and the home environment. *Journal of Applied Developmental Psychology*, *28*, 411-426, http://dx.doi.org/10.1016/j.appdev.2007.06.010

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

McCoy, D. C., Peet, E. D., Ezzati, M., Danaei, G., Black, M. M., Sudfeld, C. R., ... & Fink, G.

(2016). Early childhood developmental status in low-and middle-income countries: National, regional, and global prevalence estimates using predictive modeling. *PLoS Med*, *13*, e1002034, http://dx.doi.org/10.1371/journal.pmed.1002034

McCoy, D. C., Salhi, C., Yoshikawa, H., Britto, P., Black, M., & Fink, G. (2017). Opportunities for learning in low- and middle-income countries: A landscape analysis. *Under Review*.

Miller, E. B., Farkas, G., Vandell, D. L., & Duncan, G. J. (2014). Do the effects of Head Start vary by parental preacademic stimulation? *Child Development*, *85*(4), 1385-1400, http://dx.doi.org/10.1111/cdev.12233

Myers, R. G. (2004). In search of quality in programmes of early childhood care and education. *Background paper for Education for All, Global Monitoring Report 2005*. UNESCO.

NICHD Early Child Care Research Network. (2002). Early child care and children's development prior to school entry: Results from the NICHD Study of Early Child Care. *American Educational Research Journal*, *39*(1), 133-164, http://dx.doi.org/10.3102/00028312039001133

Nsamenang, A. B. (2006). Human ontogenesis: An indigenous African view on development and intelligence. *International Journal of Psychology*, *41*(4), 293-297, http://dx.doi.org/10.1080/00207590544000077

Nsamenang, A. B., & Lamb, M. E. (1994). Socialization of Nso children in the Bamenda grassfields of northwest Cameroon. In P. M. Greenfield & R. R. Cocking (Eds.), *Cross-Cultural Roots of Minority Child Development* (pp. 133–146). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

OECD. (2015). OECD Family Database. OECD.

Osei, G.M. (2006). Teachers in Ghana: Issues of training, remuneration and effectiveness. *International Journal of Educational Development*, *26*, 38–51, https://doi.org/10.1016/j.ijedudev.2005.07.015

Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., & Barbarin, O. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions? *Applied Developmental Science*, *9*(3), 144-159, http://dx.doi.org/10.1207/s1532480xads0903_2

Pisani, L., Borisova, I., & Dowd, A. J. (2015). *International development and early learning assessment technical working paper*. Save the Children.

Rao, N., Sun, J., Ng, M., Becher, Y., Lee, D., Ip, P., & Bacon-Shone, J. (2014). *Validation, finalization and adoption of the East Asia-Pacific Early Child Development Scales*. UNICEF.

Raver, C. C., Jones, S. M., Li-Grining, C., Zhai, F., Bub, K., & Pressler, E. (2011). CSRP's impact on low-income preschoolers' preacademic skills: Self-regulation as a mediating mechanism. *Child Development*, *82*(1), 362-378. http://dx.doi.org/10.1111/j.1467-8624.2010.01561.x

Rimm-Kaufman, S. E., Curby, T. W., Grimm, K. J., Nathanson, L., & Brock, L. L. (2009). The contribution of children's self-regulation and classroom quality to children's adaptive behaviors in the kindergarten classroom. *Developmental Psychology*, *45*(4), 958, http://dx.doi.org/10.1037/a0015861

Sandilos, L. E., DiPerna, J. C., & Family Life Project Key Investigators. (2014). Measuring quality in kindergarten classrooms: Structural analysis of the Classroom Assessment Scoring System (CLASS K–3). *Early Education and Development*, *25*(6), 894-914, https://doi.org/10.1080/10409289.2014.883588

Schreiner, M., & Woller, G. (2010). A simple poverty scorecard for Ghana. *Consultative Group to Assist the Poorest*.

Seidman, E., Raza, M., Kim, S., & McCoy, J. M. (2013). *Teacher Instructional Practices and*

*Processes System (V.5) – TIPPS: Manual and scoring system.* New York: NYU.

Serpell, R. (2011). Social responsibility as a dimension of intelligence, and as an educational goal: Insights from programmatic research in an African society. *Child Development Perspectives*, *5*(2), 126-133, http://dx.doi.org/10.1111/j.1750-8606.2011.00167.x

Squires, J., & Bricker, D. (2009). Ages & Stages Questionnaires, (ASQ-3). *A parent-completed child monitoring system. 3rd ed.* Baltimore, MD: Brookes.

Super, C. M., Harkness, S., Barry, O., & Zeitlin, M. (2011). Think locally, act globally: Contributions of African research to child development. *Child Development Perspectives, 5*(2), 119-125. http://dx.doi.org/10.1111/j.1750-8606.2011.00166.x

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 360-407, http://dx.doi.org/10.1598/RRQ.21.4.1

Thelen, E., & Smith, L. B. (1996). *A dynamic systems approach to the development of cognition and action*. MIT press.

UNESCO, UNICEF, Brookings Institution, & the World Bank (2017). *Overview MELQO: Measuring Early Learning Quality Outcomes*. Paris, France: UNESCO.

UNESCO (2015). *Education for All Global Monitoring Report Regional Overview: Sub-Saharan Africa*. Paris: UNESCO. Retrieved from: http://en.unesco.org/gem-report/sites/gem-report/files/regional_overview_SSA_en.pdf

United Nations. (2015). *Sustainable Development Goals.* New York, NY: United Nations.

Vandell, D. (2004). Early child care: The known and the unknown. *Merrill-Palmer Quarterly, 50(3)*, 387-414, http://dx.doi.org/10.1353/mpq.2004.0027

Wagner, D.A. (2018). *Learning as Development: Rethinking International Education in a Changing World.* New York, NY: Routledge.

Weiland, C., Ulvestad, K., Sachs, J., & Yoshikawa, H. (2013). Associations between classroom quality and children's vocabulary and executive function skills in an urban public prekindergarten program. *Early Childhood Research Quarterly*, *28*(2), 199-209, http://dx.doi.org/10.1016/j.ecresq.2012.12.002

Weiland, C., & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development*, *84*(6), 2112-2130, http://dx.doi.org/10.1111/cdev.12099

Wolf, S., Aber, J.L., & Behrman, J.R. (2018). Experimental Impacts of the Quality Preschool for Ghana Interventions on Teacher Professional Well-being, Classroom Quality, and Children's School Readiness. Manuscript under revision.

Wolf, S., Halpin, P., Yoshikawa, H., Pisani, L., Dowd, A., & Borisova, I. (2017). Measuring school readiness globally: Assessing the construct validity and measurement invariance of the International Development and Early Learning Assessment (IDELA) in Ethiopia. *Early Childhood Research Quarterly*, *40(4)*, 21-36.

Wolf, S., Raza, M., Kim, S., Aber, J.L, Behrman, J., & Seidman, E. (2018). Measuring classroom process quality in pre-primary classrooms in Ghana using the Teacher Instructional Practices and Processes System (TIPPS). Manuscript under revision.

World Bank. (2017). *Ghana*. Washington, DC: World Bank. Retrieved from: http://data.worldbank.org/country/ghana

Yoshikawa, H., Leyva, D., Snow, C. E., Treviño, E., Barata, M., Weiland, C., ... & Arbour, M. C. (2015). Experimental impacts of a teacher professional development program in Chile on preschool classroom quality and child outcomes. *Developmental Psychology*, *51*(3), 309. http://dx.doi.org/10.1037/a0038785

Zaslow, M., Anderson, R., Redd, Z., Wessel, J., Daneri, P., Green, K., ... & Martinez-Beck, I.

(2016). Quality thresholds, features, and dosage in early care and education: Introduction and literature review. *Monographs of the Society for Research in Child Development*, *81*(2), 7-26, http://dx.doi.org/10.1111/mono.12236

Zhang, Y. (2006). Urban-rural literacy gaps in Sub-Saharan Africa: The roles of socioeconomic status and school quality. *Comparative Education Review*, *50*(4), 581-602.

*Table 1*. Descriptive statistics for key study variables (unimputed values)

| | *n* (before imputation) | % imputed | *Mean* or % | *SD* | Min | Max |
|---|---|---|---|---|---|---|
| *Child-level characteristics (N = 3,407)* | | | | | | |
| Child gender (male) | 3,407 | 0% | 50.5% | | | |
| Child age in spring (yrs) | 3,010 | 12% | 5.82 | 1.33 | 2 | 11 |
| Caregiver age (years) | 2,633 | 23% | 38.3 | 8.88 | 17 | 80 |
| Caregiver married | 2,639 | 23% | 80.1% | | | |
| Poverty Scorecard (0-100)* | 2,640 | 23% | 59.9 | 13.93 | 6 | 92 |
| Child skills | | | | | | |
| Literacy (fall) | 2,975 | 13% | 0.46 | 0.22 | 0.00 | 0.98 |
| Literacy (spring) | 3,407 | 0% | 0.61 | 0.20 | 0.00 | 1.00 |
| Numeracy (fall) | 2,975 | 13% | 0.71 | 0.26 | 0.00 | 1.00 |
| Numeracy (spring) | 3,407 | 0% | 0.83 | 0.19 | 0.00 | 1.00 |
| Prosocial (fall) | 2,975 | 13% | 0.42 | 0.20 | 0.00 | 0.97 |
| Prosocial (spring) | 3,407 | 0% | 0.54 | 0.19 | 0.00 | 0.97 |
| Executive Function (fall) | 2,975 | 13% | 0.49 | 0.21 | 0.00 | 0.89 |
| Executive Function (spring) | 3,407 | 0% | 0.59 | 0.18 | 0.00 | 0.89 |
| Approaches to Learning (fall) | 2,974 | 13% | 3.09 | 0.75 | 1.00 | 4.00 |
| Approaches to Learning (spring) | 3,407 | 0% | 3.20 | 0.61 | 1.13 | 4.00 |
| *Classroom-level characteristics (N = 438)* | | | | | | |
| Teacher switched over year | 438 | 0% | 20.6% | | | |
| Classroom quality | | | | | | |
| Facilitate Deeper Learning (fall) | 313 | 29% | 2.12 | 0.76 | 1.00 | 4.00 |
| Facilitate Deeper Learning (spring) | 418 | 5% | 2.36 | 0.66 | 1.00 | 4.00 |
| Facilitate Deeper Learning (change) | 304 | 31% | 0.21 | 0.94 | -2.33 | 2.67 |
| Support Student Expression (fall) | 313 | 29% | 2.78 | 0.44 | 1.57 | 3.57 |
| Support Student Expression (spring) | 418 | 5% | 3.05 | 0.38 | 1.71 | 3.71 |
| Support Student Expression (change) | 304 | 31% | 0.26 | 0.54 | -1.71 | 1.71 |
| Emotional Support & Behavior Management (fall) | 313 | 29% | 1.52 | 0.59 | 1.00 | 3.50 |
| Emotional Support & Behavior Management (spring) | 418 | 5% | 1.72 | 0.65 | 1.00 | 3.50 |
| Emotional Support & Behavior Management (change) | 304 | 31% | 0.17 | 0.81 | -2.50 | 2.3 |
| Curriculum Checklist (fall) | 313 | 29% | 3.05 | 1.52 | 0.00 | 8.0 |
| Curriculum Checklist (spring) | 418 | 5% | 4.07 | 1.79 | 0.00 | 10.0 |
| Curriculum Checklist (change) | 304 | 31% | 1.00 | 2.02 | -5.00 | 6.0 |
| *School-level characteristics (N = 236)* | | | | | | |
| Private school (vs. public) | 236 | 0% | 54.2% | | | |
| District | | | | | | |
| Ga East | 236 | 0% | 16.1% | | | |
| Ga South | 236 | 0% | 24.2% | | | |
| Adenta | 236 | 0% | 13.1% | | | |
| Ledzokuku-Krower | 236 | 0% | 22.0% | | | |
| Ga Central | 236 | 0% | 14.0% | | | |
| Madina | 236 | 0% | 10.6% | | | |

*Note*: *Higher scores indicate higher levels of wealth and lower levels of poverty.

*Table 2.* Results of multi-level models predicting children's spring outcomes based on changes in classroom quality ($N$ = 3,407)

| | Literacy | | | | Numeracy | | | | Prosocial | | | | Executive Function | | | | Approaches to Learning | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *b* | | *(SE)* | *β* | *b* | | *(SE)* | *β* | *b* | | *(SE)* | *β* | *b* | | *(SE)* | *β* | *b* | | *(SE)* | *β* |
| *Classroom-Level Predictors* | | | | | | | | | | | | | | | | | | | | |
| FDL change | 0.006 | * | (0.003) | 0.030 | -0.003 | | (0.003) | -0.013 | 0.010 | ** | (0.004) | 0.052 | 0.007 | + | (0.003) | 0.035 | 0.014 | | (0.011) | 0.021 |
| ESBM change | -0.001 | | (0.005) | -0.003 | 0.012 | * | (0.005) | 0.034 | -0.013 | * | (0.006) | -0.038 | 0.001 | | (0.006) | 0.002 | -0.041 | * | (0.019) | -0.036 |
| SSE change | -0.007 | * | (0.003) | -0.028 | 0.001 | | (0.003) | 0.003 | -0.006 | + | (0.004) | -0.026 | 0.004 | | (0.004) | 0.017 | -0.022 | * | (0.011) | -0.029 |
| Curriculum checklist change | 0.006 | ** | (0.001) | 0.061 | 0.002 | * | (0.001) | 0.024 | 0.005 | ** | (0.001) | 0.057 | 0.003 | * | (0.001) | 0.037 | 0.020 | ** | (0.004) | 0.068 |
| *Child-Level Covariates* | | | | | | | | | | | | | | | | | | | | |
| Child fall outcome score | 0.466 | ** | (0.003) | | 0.565 | ** | (0.003) | | 0.332 | ** | (0.003) | | 0.305 | ** | (0.003) | | 0.252 | ** | (0.003) | |
| Child is male | -0.008 | ** | (0.001) | | 0.004 | ** | (0.001) | | -0.016 | ** | (0.001) | | -0.001 | | (0.001) | | 0.000 | | (0.004) | |
| Child age (in months) | 0.008 | ** | (0.000) | | 0.011 | ** | (0.000) | | 0.012 | ** | (0.001) | | 0.008 | ** | (0.001) | | 0.019 | ** | (0.002) | |
| Caregiver age (in years) | -0.000 | ** | (0.000) | | -0.000 | | (0.000) | | 0.000 | | (0.000) | | -0.000 | ** | (0.000) | | 0.002 | ** | (0.000) | |
| Caregiver is married | 0.003 | * | (0.001) | | 0.010 | ** | (0.001) | | -0.006 | ** | (0.002) | | -0.009 | ** | (0.001) | | -0.009 | + | (0.005) | |
| Poverty scorecard (0-100) | 0.000 | ** | (0.000) | | 0.000 | ** | (0.000) | | -0.000 | * | (0.000) | | -0.000 | ** | (0.000) | | 0.001 | ** | (0.000) | |
| *Classroom-Level Covariates* | | | | | | | | | | | | | | | | | | | | |
| Teacher switched school | -0.032 | ** | (0.007) | | -0.041 | ** | (0.006) | | -0.029 | ** | (0.008) | | -0.003 | | (0.008) | | -0.040 | + | (0.023) | |
| FDL fall | 0.010 | ** | (0.003) | | 0.001 | | (0.003) | | 0.017 | ** | (0.004) | | 0.013 | ** | (0.004) | | 0.010 | | (0.012) | |
| ESBM fall | -0.001 | | (0.006) | | 0.021 | ** | (0.005) | | -0.003 | | (0.007) | | 0.006 | | (0.007) | | -0.071 | ** | (0.020) | |
| SSE fall | 0.002 | | (0.003) | | 0.010 | ** | (0.003) | | 0.008 | * | (0.004) | | 0.005 | | (0.004) | | -0.008 | | (0.012) | |
| Curriculum checklist fall | 0.006 | ** | (0.001) | | -0.000 | | (0.001) | | -0.003 | + | (0.002) | | 0.001 | | (0.001) | | 0.021 | ** | (0.005) | |
| *School-Level Covariates* | | | | | | | | | | | | | | | | | | | | |
| School is private (vs. public) | 0.063 | ** | (0.008) | | 0.024 | ** | (0.006) | | 0.012 | | (0.008) | | 0.042 | ** | (0.008) | | 0.100 | ** | (0.025) | |
| District (ref = Ga East) | | | | | | | | | | | | | | | | | | | | |
| Ga South | 0.047 | ** | (0.012) | | 0.038 | ** | (0.010) | | 0.061 | ** | (0.012) | | 0.030 | * | (0.012) | | -0.605 | ** | (0.037) | |
| Adenta | 0.078 | ** | (0.014) | | 0.049 | ** | (0.011) | | 0.076 | ** | (0.014) | | 0.060 | ** | (0.014) | | -0.285 | ** | (0.043) | |
| Ledzokuku-Krower | -0.001 | | (0.012) | | 0.022 | * | (0.010) | | -0.001 | | (0.013) | | -0.055 | ** | (0.012) | | -0.505 | ** | (0.038) | |
| Ga Central | 0.059 | ** | (0.014) | | 0.033 | ** | (0.011) | | 0.010 | | (0.014) | | 0.014 | | (0.014) | | 0.011 | | (0.043) | |
| Madina | -0.002 | | (0.015) | | 0.002 | | (0.012) | | -0.027 | + | (0.015) | | -0.031 | * | (0.014) | | -0.544 | ** | (0.046) | |
| *Intercept* | 0.237 | ** | (0.019) | | 0.139 | ** | (0.017) | | 0.291 | ** | (0.021) | | 0.334 | ** | (0.021) | | 2.649 | ** | (0.065) | |

*Notes*: FDL = Facilitate Deeper Learning, SSE = Support Student Expression, ESBM = Emotional Support & Behavior Management; ** $p$ < .01, * $p$ < .05, + $p$ < .10. Estimates derived from 20 multiply imputed datasets.

*Table 3*. Results of multi-level models examining the interaction between fall classroom quality levels and changes in classroom quality predicting children's spring outcomes ($N$ = 3,407)

| | Literacy | | | | Numeracy | | | | Prosocial | | | | Executive Function | | | | Approaches to Learning | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *b* | | *(SE)* | *β* | *b* | | *(SE)* | *β* | *b* | | *(SE)* | *β* | *b* | | *(SE)* | *β* | *b* | | *(SE)* | *β* |
| *Interactions* | | | | | | | | | | | | | | | | | | | | |
| FDL change X FDL fall | 0.001 | | (0.001) | 0.005 | 0.002 | * | (0.001) | 0.007 | 0.004 | ** | (0.001) | 0.015 | 0.001 | | (0.001) | 0.002 | 0.007 | * | (0.004) | 0.009 |
| ESBM change X ESBM fall | 0.003 | | (0.002) | 0.004 | -0.001 | | (0.002) | -0.002 | 0.002 | | (0.003) | 0.003 | 0.012 | ** | (0.003) | 0.018 | 0.009 | | (0.009) | 0.004 |
| SSE change X SSE fall | 0.002 | | (0.002) | 0.007 | -0.003 | + | (0.001) | -0.009 | 0.002 | | (0.002) | 0.007 | -0.003 | | (0.002) | -0.010 | -0.012 | * | (0.006) | -0.012 |
| Checklist change X Checklist fall | -0.000 | * | (0.000) | -0.007 | 0.000 | | (0.000) | 0.001 | -0.000 | + | (0.000) | -0.008 | -0.001 | * | (0.000) | -0.011 | -0.001 | | (0.001) | -0.004 |
| *Classroom-Level Main Effects* | | | | | | | | | | | | | | | | | | | | |
| FDL fall | 0.010 | ** | (0.003) | 0.036 | 0.001 | | (0.003) | 0.003 | 0.017 | ** | (0.004) | 0.066 | 0.013 | ** | (0.004) | 0.053 | 0.010 | | (0.012) | 0.012 |
| ESBM fall | -0.001 | | (0.006) | -0.003 | 0.021 | ** | (0.005) | 0.050 | -0.003 | | (0.007) | -0.007 | 0.005 | | (0.007) | 0.012 | -0.072 | ** | (0.020) | -0.053 |
| SSE fall | 0.003 | | (0.004) | 0.009 | 0.009 | ** | (0.003) | 0.027 | 0.009 | * | (0.004) | 0.027 | 0.004 | | (0.004) | 0.012 | -0.012 | | (0.013) | -0.011 |
| Curriculum checklist fall | 0.006 | ** | (0.001) | 0.048 | -0.000 | | (0.001) | -0.003 | -0.003 | + | (0.002) | -0.021 | 0.001 | | (0.001) | 0.005 | 0.021 | ** | (0.005) | 0.052 |
| FDL change | 0.006 | * | (0.003) | 0.031 | -0.003 | | (0.003) | -0.016 | 0.010 | ** | (0.004) | 0.052 | 0.007 | + | (0.004) | 0.037 | 0.013 | | (0.011) | 0.021 |
| ESBM change | -0.001 | | (0.005) | -0.004 | 0.012 | * | (0.005) | 0.034 | -0.013 | * | (0.006) | -0.038 | -0.000 | | (0.006) | -0.000 | -0.041 | * | (0.019) | -0.036 |
| SSE change | -0.007 | * | (0.003) | -0.029 | 0.001 | | (0.003) | 0.003 | -0.006 | + | (0.004) | -0.027 | 0.004 | | (0.004) | 0.018 | -0.021 | + | (0.011) | -0.028 |
| Curriculum checklist change | 0.006 | ** | (0.001) | 0.062 | 0.002 | * | (0.001) | 0.024 | 0.005 | ** | (0.001) | 0.058 | 0.003 | * | (0.001) | 0.037 | 0.020 | ** | (0.004) | 0.069 |

*Notes*. FDL = Facilitate Deeper Learning, SSE = Support Student Expression, ESBM = Emotional Support & Behavior Management; ** $p < .01$, * $p < .05$, + $p < .10$; Estimates derived from 20 multiply imputed datasets; Models include full set of covariates. Coefficients for covariates not shown.

*Table 4*. Results of multi-level models examining the interaction between children's fall development scores and changes in classroom quality predicting children's spring outcomes

| | Literacy | | | | Numeracy | | | | Prosocial | | | | Executive Function | | | | Approaches to Learning | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *b* | | *(SE)* | *β* | *b* | | *(SE)* | *β* | *b* | | *(SE)* | *β* | *b* | | *(SE)* | *β* | *b* | | *(SE)* | *β* |
| *Interactions* | | | | | | | | | | | | | | | | | | | | |
| FDL change X Child Outcome fall | 0.012 | ** | (0.003) | 0.012 | 0.006 | + | (0.003) | 0.006 | 0.015 | ** | (0.004) | 0.015 | -0.019 | ** | (0.004) | -0.020 | 0.022 | ** | (0.003) | 0.025 |
| ESBM change X Child Outcome fall | 0.004 | | (0.005) | 0.003 | -0.017 | ** | (0.006) | -0.009 | -0.037 | ** | (0.007) | -0.021 | -0.032 | ** | (0.006) | -0.019 | -0.018 | ** | (0.005) | -0.012 |
| SSE change X Child Outcome fall | 0.019 | ** | (0.004) | 0.016 | 0.034 | ** | (0.004) | 0.028 | 0.022 | ** | (0.004) | 0.018 | 0.022 | ** | (0.004) | 0.020 | -0.018 | ** | (0.004) | -0.017 |
| Checklist change X Child Outcome fall | -0.000 | | (0.001) | -0.001 | 0.003 | * | (0.002) | 0.007 | 0.011 | ** | (0.002) | 0.024 | -0.006 | ** | (0.002) | -0.014 | 0.004 | ** | (0.001) | 0.011 |
| *Classroom-Level Main Effects* | | | | | | | | | | | | | | | | | | | | |
| FDL change | 0.006 | * | (0.003) | 0.030 | -0.003 | | (0.003) | -0.015 | 0.010 | ** | (0.004) | 0.054 | 0.007 | * | (0.004) | 0.040 | 0.012 | | (0.011) | 0.020 |
| ESBM change | -0.001 | | (0.005) | -0.003 | 0.012 | * | (0.005) | 0.035 | -0.013 | * | (0.006) | -0.037 | 0.001 | | (0.006) | 0.004 | -0.039 | * | (0.019) | -0.034 |
| SSE change | -0.008 | * | (0.003) | -0.031 | 0.000 | | (0.003) | 0.000 | -0.006 | + | (0.004) | -0.026 | 0.003 | | (0.004) | 0.015 | -0.020 | + | (0.011) | -0.026 |
| Curriculum checklist change | 0.006 | ** | (0.001) | 0.063 | 0.002 | * | (0.001) | 0.025 | 0.005 | ** | (0.001) | 0.057 | 0.003 | * | (0.001) | 0.039 | 0.019 | ** | (0.004) | 0.066 |
| *Child-Level Main Effects* | | | | | | | | | | | | | | | | | | | | |
| Lagged outcome (fall) | 0.466 | ** | (0.003) | 0.503 | 0.564 | ** | (0.003) | 0.583 | 0.331 | ** | (0.003) | 0.350 | 0.306 | ** | (0.003) | 0.346 | 0.252 | ** | (0.003) | 0.306 |

*Notes:* FDL = Facilitate Deeper Learning, SSE = Support Student Expression, ESBM = Emotional Support & Behavior Management; ** $p < .01$, * $p < .05$, + $p < .10$; Estimates derived from 20 multiply imputed datasets; Models include full set of covariates. Coefficients for covariates not shown.

*Appendix Table 1*. Correlations between study variables.

| | | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) | (15) | (16) | (17) | (18) | (19) | (20) | (21) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (1) | FDL fall | 1.00 | | | | | | | | | | | | | | | | | | | | |
| (2) | FDL spring | 0.13 | 1.00 | | | | | | | | | | | | | | | | | | | |
| (3) | FDL change | -0.71 | 0.60 | 1.00 | | | | | | | | | | | | | | | | | | |
| (4) | ESBM fall | 0.38 | 0.15 | -0.19 | 1.00 | | | | | | | | | | | | | | | | | |
| (5) | ESBM spring | 0.13 | 0.33 | 0.13 | 0.17 | 1.00 | | | | | | | | | | | | | | | | |
| (6) | ESBM change | -0.21 | 0.12 | 0.25 | -0.70 | 0.58 | 1.00 | | | | | | | | | | | | | | | |
| (7) | SSE fall | 0.40 | 0.18 | -0.21 | 0.14 | 0.08 | -0.05 | 1.00 | | | | | | | | | | | | | | |
| (8) | SSE spring | 0.10 | 0.40 | 0.20 | 0.03 | 0.17 | 0.12 | 0.10 | 1.00 | | | | | | | | | | | | | |
| (9) | SSE change | -0.22 | 0.17 | 0.30 | -0.07 | 0.10 | 0.13 | -0.64 | 0.71 | 1.00 | | | | | | | | | | | | |
| (10) | Curriculum checklist fall | 0.20 | 0.15 | -0.05 | 0.38 | 0.13 | -0.21 | 0.33 | 0.00 | -0.24 | 1.00 | | | | | | | | | | | |
| (11) | Curriculum checklist spring | 0.00 | 0.25 | 0.14 | 0.08 | 0.26 | 0.10 | 0.13 | 0.36 | 0.20 | 0.20 | 1.00 | | | | | | | | | | |
| (12) | Curriculum checklist change | -0.14 | 0.05 | 0.15 | -0.21 | 0.09 | 0.24 | -0.14 | 0.32 | 0.35 | -0.58 | 0.68 | 1.00 | | | | | | | | | |
| (13) | Literacy fall | 0.13 | 0.05 | -0.08 | 0.06 | 0.06 | 0.01 | 0.11 | 0.01 | -0.07 | -0.05 | -0.09 | -0.04 | 1.00 | | | | | | | | |
| (14) | Literacy spring | 0.14 | 0.10 | -0.05 | 0.06 | 0.10 | 0.03 | 0.14 | 0.05 | -0.07 | -0.03 | -0.03 | 0.01 | 0.69 | 1.00 | | | | | | | |
| (15) | Numeracy fall | 0.12 | 0.07 | -0.05 | 0.04 | 0.06 | 0.03 | 0.14 | 0.02 | -0.07 | -0.02 | -0.03 | -0.01 | 0.75 | 0.67 | 1.00 | | | | | | |
| (16) | Numeracy spring | 0.13 | 0.08 | -0.06 | 0.05 | 0.08 | 0.04 | 0.14 | 0.04 | -0.07 | 0.00 | 0.00 | 0.01 | 0.64 | 0.74 | 0.73 | 1.00 | | | | | |
| (17) | Prosocial fall | 0.08 | 0.02 | -0.06 | 0.06 | 0.01 | -0.02 | 0.07 | -0.01 | -0.07 | -0.03 | -0.01 | -0.01 | 0.55 | 0.40 | 0.48 | 0.41 | 1.00 | | | | |
| (18) | Prosocial spring | 0.10 | 0.07 | -0.03 | 0.06 | 0.04 | 0.00 | 0.12 | 0.04 | -0.08 | -0.02 | 0.03 | 0.03 | 0.42 | 0.51 | 0.40 | 0.47 | 0.45 | 1.00 | | | |
| (19) | Executive function fall | 0.07 | 0.01 | -0.06 | 0.00 | 0.01 | 0.03 | 0.09 | -0.03 | -0.10 | -0.03 | -0.05 | -0.01 | 0.50 | 0.43 | 0.52 | 0.47 | 0.43 | 0.30 | 1.00 | | |
| (20) | Executive function spring | 0.08 | 0.04 | -0.05 | 0.04 | 0.07 | 0.01 | 0.09 | -0.01 | -0.09 | -0.01 | -0.05 | -0.01 | 0.44 | 0.53 | 0.47 | 0.54 | 0.29 | 0.43 | 0.43 | 1.00 | |
| (21) | Approaches to learning fall | 0.08 | 0.03 | -0.04 | 0.03 | 0.02 | 0.02 | 0.10 | 0.01 | -0.07 | 0.00 | -0.05 | -0.05 | 0.51 | 0.41 | 0.51 | 0.43 | 0.43 | 0.29 | 0.40 | 0.33 | 1.00 |
| (22) | Approaches to learning spring | -0.01 | -0.04 | -0.04 | 0.02 | -0.01 | -0.01 | 0.07 | -0.07 | -0.09 | -0.01 | 0.01 | 0.04 | 0.44 | 0.50 | 0.44 | 0.42 | 0.32 | 0.35 | 0.33 | 0.38 | 0.40 |

*Appendix Table 2*. Results of sensitivity analyses.

| | Literacy | | | | Numeracy | | | | Prosocial | | | | Executive Function | | | | Approaches to Learning | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *b* | | *(SE)* | *β* | *b* | | *(SE)* | *β* | *b* | | *(SE)* | *β* | *b* | | *(SE)* | *β* | *b* | | *(SE)* | *β* |
| *Primary results (N = 3,407)* | | | | | | | | | | | | | | | | | | | | |
| FDL change | 0.006 | * | (0.003) | 0.030 | -0.003 | | (0.003) | -0.013 | 0.010 | ** | (0.004) | 0.052 | 0.007 | + | (0.003) | 0.035 | 0.014 | | (0.011) | 0.021 |
| ESBM change | -0.001 | | (0.005) | -0.003 | 0.012 | * | (0.005) | 0.034 | -0.013 | * | (0.006) | -0.038 | 0.001 | | (0.006) | 0.002 | -0.041 | * | (0.019) | -0.036 |
| SSE change | -0.007 | * | (0.003) | -0.028 | 0.001 | | (0.003) | 0.003 | -0.006 | + | (0.004) | -0.026 | 0.004 | | (0.004) | 0.017 | -0.022 | * | (0.011) | -0.029 |
| Curriculum checklist change | 0.006 | ** | (0.001) | 0.061 | 0.002 | * | (0.001) | 0.024 | 0.005 | ** | (0.001) | 0.057 | 0.003 | * | (0.001) | 0.037 | 0.020 | ** | (0.004) | 0.068 |
| *Fixed effects models (N = 3,407)* | | | | | | | | | | | | | | | | | | | | |
| FDL change | 0.016 | ** | (0.006) | 0.073 | 0.000 | | (0.004) | 0.002 | 0.005 | | (0.005) | 0.026 | 0.012 | * | (0.006) | 0.063 | 0.021 | | (0.015) | 0.032 |
| ESBM change | 0.006 | | (0.008) | 0.016 | 0.020 | ** | (0.007) | 0.056 | -0.006 | | (0.010) | -0017 | -0.011 | | (0.010) | -0.032 | 0.028 | | (0.021) | 0.025 |
| SSE change | -0.003 | | (0.005) | -0.012 | -0.006 | | (0.004) | -0.027 | -0.001 | | (0.006) | -0.002 | -0.010 | + | (0.005) | -0.045 | -0.013 | | (0.015) | -0.017 |
| Curriculum checklist change | 0.002 | | (0.002) | 0.020 | -0.000 | | (0.002) | -0.000 | 0.003 | | (0.002) | 0.032 | 0.004 | + | (0.002) | 0.048 | 0.003 | | (0.006) | 0.009 |
| *With listwise deletion (N = 1,698)* | | | | | | | | | | | | | | | | | | | | |
| FDL change | 0.012 | + | (0.007) | 0.058 | -0.002 | | (0.006) | -0.009 | 0.008 | | (0.008) | 0.038 | 0.007 | | (0.008) | 0.033 | -0.020 | | (0.024) | -0.030 |
| ESBM change | 0.007 | | (0.013) | 0.018 | 0.015 | | (0.010) | 0.044 | -0.009 | | (0.013) | -0.025 | -0.010 | | (0.014) | -0.028 | 0.026 | | (0.040) | 0.023 |
| SSE change | -0.002 | | (0.008) | -0.009 | -0.004 | | (0.007) | -0.018 | -0.011 | | (0.009) | -0.047 | -0.014 | | (0.009) | -0.060 | 0.011 | | (0.026) | 0.015 |
| Curriculum checklist change | 0.007 | * | (0.003) | 0.075 | 0.005 | + | (0.003) | 0.051 | 0.007 | * | (0.003) | 0.074 | 0.008 | * | (0.004) | 0.086 | 0.026 | ** | (0.010) | 0.086 |
| *Control group only (N = 933)* | | | | | | | | | | | | | | | | | | | | |
| FDL change | 0.008 | + | (0.004) | 0.035 | -0.002 | | (0.004) | -0.012 | 0.011 | * | (0.005) | 0.055 | 0.002 | | (0.005) | 0.011 | -0.005 | | (0.016) | -0.007 |
| ESBM change | -0.007 | | (0.008) | -0.020 | 0.002 | | (0.007) | 0.006 | -0.024 | ** | (0.009) | -0.068 | -0.005 | | (0.009) | -0.015 | -0.047 | | (0.029) | -0.041 |
| SSE change | 0.001 | | (0.005) | 0.003 | 0.004 | | (0.005) | 0.017 | -0.012 | * | (0.006) | -0.053 | -0.011 | + | (0.006) | -0.050 | 0.000 | | (0.019) | 0.000 |
| Curriculum checklist change | 0.006 | ** | (0.002) | 0.065 | 0.005 | ** | (0.002) | 0.050 | 0.009 | ** | (0.002) | 0.095 | 0.003 | | (0.002) | 0.032 | 0.028 | ** | (0.007) | 0.094 |
| *Treatment groups only (N = 2,042)* | | | | | | | | | | | | | | | | | | | | |
| FDL change | 0.017 | ** | (0.005) | 0.080 | -0.002 | | (0.004) | -0.011 | 0.005 | | (0.006) | 0.024 | 0.008 | | (0.006) | 0.042 | 0.007 | | (0.017) | 0.010 |
| ESBM change | -0.010 | | (0.009) | -0.026 | 0.013 | + | (0.008) | 0.037 | -0.012 | | (0.010) | -0.034 | -0.010 | | (0.010) | -0.030 | -0.060 | * | (0.030) | -0.052 |
| SSE change | -0.021 | ** | (0.004) | -0.084 | -0.010 | ** | (0.004) | -0.043 | -0.005 | | (0.005) | -0.021 | 0.014 | ** | (0.005) | 0.060 | -0.025 | + | (0.015) | -0.033 |
| Curriculum checklist change | 0.006 | ** | (0.002) | 0.065 | 0.001 | | (0.002) | 0.013 | 0.003 | | (0.002) | 0.031 | 0.005 | * | (0.002) | 0.059 | 0.015 | * | (0.006) | 0.052 |
| *Excluding classrooms with teacher turnover (N = 2,700)* | | | | | | | | | | | | | | | | | | | | |
| FDL change | 0.004 | | (0.003) | 0.020 | -0.005 | + | (0.003) | -0.026 | 0.007 | + | (0.004) | 0.035 | 0.005 | | (0.004) | 0.026 | 0.010 | | (0.012) | 0.016 |
| ESBM change | 0.006 | | (0.006) | 0.017 | 0.014 | * | (0.006) | 0.039 | -0.016 | * | (0.007) | -0.046 | 0.000 | | (0.007) | 0.001 | -0.041 | + | (0.022) | -0.036 |
| SSE change | 0.002 | | (0.004) | 0.009 | 0.002 | | (0.003) | 0.010 | 0.002 | | (0.004) | 0.010 | -0.007 | | (0.004) | -0.030 | -0.016 | | (0.013) | -0.021 |
| Curriculum checklist change | 0.005 | ** | (0.001) | 0.052 | 0.001 | | (0.001) | 0.008 | 0.004 | * | (0.002) | 0.042 | 0.004 | * | (0.002) | 0.040 | 0.019 | ** | (0.005) | 0.064 |
| *Each domain individually (N = 3,407)* | | | | | | | | | | | | | | | | | | | | |
| FDL change | 0.008 | ** | (0.003) | 0.038 | 0.002 | | (0.003) | 0.011 | 0.010 | ** | (0.003) | 0.051 | 0.011 | ** | (0.003) | 0.053 | 0.011 | | (0.010) | 0.017 |
| ESBM change | 0.010 | + | (0.005) | 0.026 | 0.014 | ** | (0.005) | 0.040 | -0.003 | | (0.006) | -0.007 | 0.011 | + | (0.006) | 0.031 | -0.010 | | (0.018) | -0.008 |
| SSE change | 0.001 | | (0.003) | 0.004 | 0.004 | | (0.002) | 0.016 | 0.003 | | (0.003) | 0.012 | 0.010 | ** | (0.003) | 0.045 | -0.002 | | (0.010) | -0.002 |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Curriculum checklist change | 0.006 | ** | (0.001) | 0.058 | 0.003 | ** | (0.001) | 0.035 | 0.005 | ** | (0.001) | 0.053 | 0.005 | ** | (0.001) | 0.053 | 0.015 | ** | (0.004) | 0.052 |

*Notes*. FDL = Facilitate Deeper Learning, SSE = Support Student Expression, ESBM = Emotional Support & Behavior Management; ** $p < .01$, * $p < .05$, + $p < .10$; Primary results reflect results shown in Table 2. "Fixed effects models" indicate results of models that include school fixed effects and clustering at the classroom level rather than multi-level models. "With listwise deletion" indicates results of models in which missing values were addressed through listwise deletion rather than multiple imputation. "Control group only" indicates results of analyses run in the control group of the parent study's intervention trial. "Treatment groups only" indicates results of analyses run in the treatment groups of the parent study's intervention trial. "Excluding classrooms with teacher turnover" indicates results of analyses run in subgroup of children whose teachers were the same from fall to spring. "Each domain individually" indicates results of models in which each dimension of classroom quality was entered separately; All models include full set of covariates. Coefficients for covariates not shown.