GSE Faculty Research   Graduate School of Education

10-28-2019

Year 3 Follow-up of the 'Quality Preschool for Ghana' Interventions on Child Development

Sharon Wolf
University of Pennsylvania, wolfs@upenn.edu

Year 3 Follow-up of the 'Quality Preschool for Ghana' Interventions on Child Development

Abstract

Governments around the world are increasing investments in early childhood education as a way to promote children's learning and development. As research grows on the longer-term effects of early educational programs, some have hypothesized that sustained impacts may depend on the quality of children's subsequent classroom environments and may be more likely to occur in certain nonacademic domains. This study is the first to examine these questions in sub-Saharan Africa, using longitudinal data from a school-randomized trial in Ghana. A 1-year teacher-training and coaching program, delivered with (TTPA) and without (TT) parental-awareness meetings, was implemented when children were enrolled in 1 of 2 grades of preprimary school. Previous studies showed that in Year 1, the implementation year, the TT treatment had positive impacts on children's literacy, numeracy, and social-emotional development, while the TTPA treatment yielded no positive impacts for children in both grades. In this study, I examine impacts in Year 3, at the end of children's 1st or 2nd year of primary school. I find persistent positive impacts of TT on literacy, and new negative impacts of TTPA on numeracy, but these depend on levels of classroom emotional support and teacher burnout in primary school. In addition, there were small persistent impacts of both treatments on children's executive function, and these were not conditional on subsequent classroom environments.

Keywords

early childhood education, preschool, sub-Saharan Africa, persistence, convergence, randomized control trial

Disciplines

Developmental Psychology | Early Childhood Education | Education | International and Comparative Education | Pre-Elementary, Early Childhood, Kindergarten Teacher Education | Teacher Education and Professional Development

Year 3 Follow-up of the 'Quality Preschool for Ghana' Interventions on Child

Development

Manuscript accepted for publication

Abstract

Governments around the world are increasing investments in early childhood education as a way to promote children's learning and development. As research grows on the longer-term effects of early educational programs, some have hypothesized that sustained impacts may depend on the quality of children's subsequent classroom environments and may be more likely to occur in certain nonacademic domains. This study is the first to examine these questions in sub-Saharan Africa, using longitudinal data from a school-randomized trial in Ghana. A 1-year teacher-training and coaching program, delivered with (TTPA) and without (TT) parental-awareness meetings, was implemented when children were enrolled in one of two grades of pre-primary school. Previous studies showed that in Year 1, the implementation year, the TT treatment had positive impacts on children's literacy, numeracy, and social-emotional development, while the TTPA treatment yielded no positive impacts for children in both grades. In this study, I examine impacts in Year 3, at the end of children's first or second year of primary school. I find persistent positive impacts of TT on literacy, and new negative impacts of TTPA on numeracy, but these depend on levels of classroom emotional support and teacher burnout in primary school. In addition, there were small persistent impacts of both treatments on children's executive function, and these were not conditional on subsequent classroom environments.

Keywords: early childhood education; preschool; sub-Saharan Africa; persistence; convergence; randomized control trial

The preschool years are a period of rapid growth in cognitive, social, and emotional skills, and a sensitive period for children's educational and developmental potential (Shonkoff & Philips, 2000). Vast inequalities exist in young children's learning opportunities and outcomes around the world, and there is growing interest in expanding access to early childhood education (ECE) as a mechanism to reduce inequality both within (Phillips et al., 2017) and across countries (United Nations, 2015; Yoshikawa, Wuermli, Raikes, Kim, & Kabay., 2018). Governments in low- and middle-income countries (LMICs) have been increasing investments in ECE programs (Engle et al., 2011; Yoshikawa & Kabay, 2014). Yet as programs have expanded, concerns have arisen about their quality, including whether ECE services are structured to promote children's learning and satisfy parents' desires for education (Yoshikawa et al., 2018). Thus, it remains unclear whether the expansion of ECE services in LMICs is improving children's development and learning.

ECE programs that have more lasting effects and improve primary school learning levels will be more cost-effective compared with those for which impacts converge after exposure to the program ends. The longer-term persistence or convergence of gains from ECE has received attention in the United States (e.g., Bailey, Duncan, Odgers, & Yu, 2017; Jenkins et al., 2018), but not yet in LMICs. A recent review of randomized control trials in poor countries found that very few studies have evaluated longer-term program impacts, and the authors highlighted a need for high-quality evidence on the long-run impacts of interventions in LMICs (Bouguen, Huang, Kremer, & Miguel, 2019). In parts of the world where children face extreme levels of risk, there is great potential for ECE to support children's long-term development. 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 children age 3 and 4 (29.4 million; 44%) failing to meet cognitive and social-emotional milestones (McCoy et al., 2016). At the same time, compared to other regions, SSA has the lowest ECE enrollment rates, at

around 18% (McCoy et al., 2017), and learning levels are extremely low for primary school students (Sandefur, 2016). On the one hand, in this context ECE could boost children's school readiness and ultimately improve learning trajectories in primary school. On the other hand, if subsequent educational environments are poor, there may be little chance for sustained gains.

The present study examines three critical questions about ECE and global child development, using longitudinal data from a school-randomized trial in Ghana (Wolf, Aber, Behrman, & Tsinigo, 2019). First, does an intervention previously shown to improve ECE quality and produce positive short-run effects (i.e., over the implementation year and 1 year later) also show sustained gains for children in the third year, when children are in primary school? And, does a second intervention engaging parents that had no positive effects also show sustained null impacts? Second, do the quality of children's primary school classrooms and the level of primary school teachers' burnout moderate sustained gains? This is the first study to examine how subsequent environments moderate sustained impacts of exposure to quality ECE in SSA.

Foundational Skills for Long-term Impacts

As children's neurodevelopmental capacity for higher-order thinking increases during the preschool years (Shonkoff & Phillips, 2000; Zelazo & Carlson, 2012), exposure to quality ECE has the potential to improve school readiness across multiple domains in ways that may lead to longer-term impacts (Snow & Van Hemel, 2008). The domain-specific skills of early literacy and numeracy are powerful predictors of later academic achievement because they form the foundation for acquiring higher-level academic skills (Duncan et al., 2007). Social-emotional (SE) and executive function (EF) skills are domain-general nonacademic constructs that have also been identified as important components of children's school success (Durlak et al., 2011; Raver, 2003). Higher-order cognitive processes such as EF help children control impulses, maintain and shift attention, and manipulate information in working memory (Blair, 2002; Miyake et al., 2000). SE

skills generally include the abilities to recognize and manage emotions, appreciate the perspectives of others, constructively handle interpersonal conflicts, and make responsible decisions (CASEL, 2017). Such prosocial skills can foster positive relationships, emotional competencies, and social problem-solving skills (e.g., Curby, Brown, Bassett, & Denham, 2015).

A broad body of correlational research suggests that children's early EF skills are predictive of later academic outcomes (e.g., Bull & Lee, 2014). Correlational research on the links between early SE skills and later learning outcomes is more mixed, with numerous studies showing positive associations (e.g., Arnold, Kupersmidt, Voegler-Lee, & Marshall, 2012; Curby et al., 2015), and others showing no such links (e.g., Duncan et al., 2007). Experimental research provides additional evidence that nonacademic skills may play a key role in improving children's learning, with several randomized control trials of SE and EF interventions showing impacts on children's language, literacy, and math outcomes, despite the fact that these academic skills were not directly targeted (Durlak et al., 2011).

Emerging evidence also suggests that these relations are not unidirectional. Instead, it is also possible that children's early growth in academic skills predicts their development of SE and EF skills. In a recent review, Clements, Sarama, and Germeroth (2016) posited that the development of early math skills may also translate into gains in children's skills in EF, as learning math "exercises" children's working memory and cognitive flexibility, both of which are central to EF. Experimental examples show that participation in a preschool math curriculum reduced behavior problems (Dobbs, Doctoroff, Fisher, & Arnold, 2006), whereas implementation of a language and mathematics curricula improved multiple dimensions of EF (Weiland & Yoshikawa, 2013). In sum, improving domains in children's early academic skills in preschool may have longer-term effects on nonacademic skills, and vice versa.

Examining Persistence and Convergence in ECE Programs

One key motivation for examining sustained effects of quality ECE stems from a few prominent studies in the United States showing strikingly positive impacts in adulthood and large benefit-cost ratios of small-scale but intensive interventions, including the Perry Preschool (e.g., Heckman, Moon, Pinto, Savelyev, & Yavitz, 2010) and Abecedarian studies (e.g., Campbell et al., 2012). The broader experimental literature analyzing longer-term impacts of ECE in high-income countries has shown mixed results for persistent impacts in the years following interventions. Some studies document small, continued benefits into middle childhood and even adolescence (Ansari 2018; Dodge, Bai, Ladd, & Muschkin, 2017; Gormley, Phillips, & Anderson, 2018; Vandell, Burchinal, & Pierce, 2016), and others find no detectable effects shortly after kindergarten (Hill, Gormley, & Adelstein, 2015; Magnuson, Ruhm, & Waldfogel, 2007; Puma et al., 2012). In a recent meta-analysis, McCoy and colleagues (2017) found strong support for sustained gains on educational outcomes into adolescence, including reductions in grade retention and increases in high school graduation rates. Some studies suggest that ECE may have long-term benefits into adulthood by improving children's nonacademic skills, such as self-regulation and executive function (e.g., Blair & Raver, 2014; Diamond, Barnett, Thomas, & Munro, 2007). Taken together, the findings suggest that the potential benefits of high-quality ECE may be improved social and executive function skills in the short-term that may translate in to improved academic outcomes in the longer term.

Sustaining Environments

There is some evidence that children's subsequent classroom and school contexts matter for whether gains from preschool last into elementary school, what some have called "sustaining environments" (Bailey et al., 2017). There are limited studies testing this hypothesis, and variation in both methodological designs and operationalization of subsequent educational quality makes it difficult to draw conclusions across the studies. For example, evidence that sustained gains are

larger in higher quality subsequent school and classroom contexts has been found for school spending (Johnson, 2013), teacher value-added (Swain, Springer, & Hofer, 2015), school-level standardized test scores (Currie & Thomas, 2000; Zhai, Raver, & Jones, 2012), alignment of preschool and kindergarten experiences (Clements, Sarama, Wolfe, & Spitler, 2013), safety and academic environment (Lee & Loeb, 1995), and a multidimensional measure of school quality including school resources, organization of school resources, and social and relational processes (Ansari & Pianta, 2018). Furthermore, using two classroom quality measures—time spent on instruction and class size—Magnuson and colleagues (2007) found that advantages from attending preschool persisted in *lower* quality classrooms, and that children who did not attend preschool "caught up" in higher quality classrooms. In contrast, others have found that the long-term benefits of early education programs are not conditional on the quality of children's later schooling and, ultimately, the benefits of these programs diminish to a similar degree for all children as they progress through elementary schooling. For example, Bassok, Gibbs, and Latham (2015) found no interaction between preschool experiences and kindergarten-year quality measures, including class size, peer preschool experience, full-day kindergarten, and an index for quality.

The broader literature on classroom quality points to process quality—the nature of children's daily interactions and experiences in the classroom, with a focus on the social, emotional, physical, and instructional aspects of activities and interactions—as the driver of child development and learning (Howes, James, & Ritchie, 2003; Pianta et al., 2005). There are three critical domains of process quality: instructional support, emotional support, and classroom organization, with instructional and emotional support shown to be most strongly predictive of child development (Hamre et al., 2013). While elements of process quality have been examined in studies testing the sustaining environments hypothesis (e.g., Ansari & Pianta, 2018), these specific domains have not been directly tested. It is possible that the inconsistencies found in the literature

are related to differential correlations between the indicators used and children's daily experiences in the classroom with their teachers (Pianta et al., 2005). Some indicators may be more highly correlated with classroom process quality (e.g., social and relational processes) than others (e.g., full- vs. half-day kindergarten).

Also yet to be considered as a feature of sustaining environments is teachers' professional well-being, despite widespread recognition that in many countries teachers face several professional challenges that can lead to stress and burnout. In LMICs in particular, there is concern that teachers are increasingly demotivated and burned out, which may partially explain deteriorating teaching performance and student learning outcomes, high rates of turnover and absenteeism, and misconduct (Bennell & Akyeampong, 2007; Moon, 2007). Burnout may be due to education reforms in the 1990s that led to a rapid rise in the number of children enrolled in school, increasing class sizes and teachers' workloads without sufficient support (VSO, 2002). Emotional exhaustion—including the depletion of emotional resources and a feeling of no longer being able to give of oneself at a psychological level—is the central element of burnout (Maslach, Jackson, & Leiter, 1996). Burnout can lead to negative, cynical attitudes and feelings about ones' students or colleagues (Skaalvik & Skaalvik, 2010), and a recent study found it to be directly associated with lower student achievement (McLean & Connor, 2015). Even if teachers teach effectively in the classroom, burnout may impact interactions with students throughout the day, including outside of instructional time, ultimately affecting how students engage in the classroom. Many have argued that teacher professional well-being generally, and burnout particularly, is an important dimension of the classroom context worthy of investigation in and of itself (Wolf & Lee, 2019; Wolf, Aber, Torrente, McCoy, & Rasheed, 2015).

Evaluations of ECE in Low- and Middle-Income Countries

Over the past 15 years, there has been a rapid expansion of ECE services around the world

(UNESCO, 2015). However, enrollment and quality of services have been generally lower in LMICs (Neuman, Josephson, & Chua, 2015). Studies in LMICs have mostly focused on short-run impacts, typically within a year of program initiation. In Chile, Yoshikawa and colleagues (2015) found that although a teacher workshop and coaching intervention had positive impacts on classroom quality, it did not translate into gains for child development outcomes over 2 school years of implementation. Araujo and colleagues (2016) randomly assigned Ecuadorian children to kindergarten teachers across 204 schools. A one-standard-deviation increment in classroom quality (total CLASS scores) predicted increases in children's language, math, and executive function of 0.11, 0.11, and 0.07 standard deviations, respectively, over 1 school year. In Bangladesh, children exposed to a high-quality preschool program outperformed a control group in verbal and nonverbal reasoning (Aboud, 2006).

A few studies have examined longer-term outcomes. In Argentina, access to 1 year of pre-primary school increased third grade language and mathematics scores by 0.23 standard deviations and had positive effects on self-control (Berlinski, Galiani, & Gertler, 2009). In Mozambique, in comparison to a control group, children at a rural preschool were 24% more likely to enroll in primary school and showed improved cognitive abilities, fine-motor skills, and behavioral outcomes (Martinez, Naudeau, & Pereira, 2012). Finally, a recent experimental study in Malawi found that a 1-year teacher-training program delivered with a parenting-skills intervention improved children's language skills and prosocial behaviors by 0.2 to 0.3 standard deviations at an 18-month follow-up, but there were no detectable impacts at the 36-month follow-up (Ozler et al., 2018). Taken together, the limited evidence suggests that across diverse LMIC contexts, similar to evidence from the United States, quality pre-primary education has short-term impacts on child development across a broad set of outcomes, with mixed evidence on longer-term impacts.

The Quality Preschool for Ghana (QP4G) Programs

The Republic of Ghana is a unique context to study the issues of early childhood education. In 2007, 2 years of pre-primary education—called *Kindergarten 1* (KG1; equivalent to pre-K in the United States) and *Kindergarten 2* (KG2; equivalent to U.S. kindergarten), respectively—were added to the universal basic education system that had previously begun in primary Grade 1. Ghana has among the highest net enrollment rates in Africa (UNESCO, 2015), with 75% of 4- and 5-year-olds enrolled in pre-primary education (Ghana Ministry of Education, 2016). But several reports have concluded that classroom quality and learning outcomes are low (e.g., Ghana Ministry of Education, 2014), including in kindergarten (Ghana Education Service, 2012). A 2012 government Kindergarten Situational Report concluded that the curriculum established in 2004 was sound, but that teachers had not been systematically trained to deliver it (Ghana Education Service, 2012). The QP4G project aimed to train teachers on the 2004 KG curriculum to enhance the quality of KG education and children's developmental outcomes, and to test the added benefits of engaging parents through an awareness campaign designed to align parental expectations with the KG curriculum and pedagogy. *Early childhood education* and *preschool* in this study refer to the 2 years of pre-primary education in Ghana called *kindergarten*.

The programs. There were two components of the QP4G interventions, which are described in detail elsewhere (Wolf et al., 2019). The primary component targeted teachers through in-service training and coaching. The training workshops (5 days in September, 2 days in January, 1 day in May) were led by professional teacher trainers at the National Nursery Teacher Training Center (NNTTC) in Accra, and in-classroom coaching (six visits over the school year) were administered by trained district-government ECE coordinators. The coaching model has a strong evidence base in both the United States (Kraft, Blazar, & Hogan, 2018) and LMICs (Ganimian & Murnane, 2016).

The second component targeted parents, with the goal of aligning expectations for ECE with the new methods taught in the training and encouraging communication between parents and teachers. The program consisted of three parental-awareness meetings, which were administered through school Parent–Teacher Associations (PTAs) over the school year. They were offered to all parents with kindergarten children in the school and administered by the same trained district-government ECE coordinators. Each meeting consisted of viewing videos developed for the intervention followed by discussions led by district coordinators. The video themes were (a) the importance of play-based learning, (b) parents' role in children's learning, and (c) encouraging parent–teacher and parent–school communication.

Mechanisms of change. Treatment schools received the teacher-training and coaching program (TT) or the teacher-training and coaching program plus the parental-awareness meetings (TTPA). The broad theory of change was that children's school readiness would be enhanced through improved classroom quality (measured through teacher–child interactions) and improved teacher professional well-being (measured through teachers' motivation, burnout, and job satisfaction). For schools that received both teacher-training and parental-awareness meetings (TTPA), it was anticipated that this combined package would improve parents' understanding of play-based learning, strengthen parent–teacher communication, and increase parental involvement and engagement with teachers, and consequently have larger impacts on child outcomes.

The most proximal target of change was the classroom, with improvements that could support children's development of foundational skills. The content focused on integrating play- and activity-based, child-centered teaching practices into instructional content. The first target was to transform classroom instruction from rote memorization of academic concepts to an activity-based approach that engages students in developing critical thinking skills. The second was to reduce harsh and corporal punishment while improving positive behavior management in the

classroom. The goal was to enhance pathways for learning by increasing emotional support and proactive behavior management to help children learn to self-regulate and develop executive function and social-emotional skills (e.g., Diamond et al., 2007).

It is important to note that perceived goals of education likely affect how teachers shape their classrooms and how they respond to parents' preferences in ways that could also affect sustained gains. Many Ghanaian parents' value early education and demand academically focused, rigorous instruction from teachers (Bidwell, Parry, & Watine, 2014; Kabay, Wolf, & Yoshikawa, 2017). Increasing parent participation and engagement in education without changing parent preferences may lead these demands to more strongly influence teachers. As parents' engagement in school increases, they may also begin to feel more efficacious and take a more active role in home learning activities that could have longer-term impacts on children. Virtually no research has examined these issues.

Previous findings. Two studies have reported on earlier impacts of the QP4G programs. The initial study tested the effectiveness of the programs, guided by the theory of change, to improve (a) teacher professional well-being, (b) classroom quality, and (c) children's developmental skills over 1 school year (Wolf et al., 2019). There were reductions in teacher burnout in TT (d_{wt}[1]= -.40) and TTPA (d_{wt} = -0.59) as well as a 58% drop in the odds of teacher turnover. Regarding classroom quality, QP4G increased the number of developmentally appropriate activities teachers used in classrooms in both treatment conditions by similar magnitudes (d_{wt} = 0.56 in TT, 0.61 in TTPA). Both treatment conditions increased levels of emotional support and positive behavior management observed in the classroom (d_{wt}= 0.65 in TT,

---

[1] d_{wT} represents a standardized mean difference between treatment and control schools accounting for nesting of children within classrooms and classrooms within schools, as calculated following Hedges (2009).

0.66 in TTPA), and TT (but not TTPA) increased levels of supporting student expression (d_{wt} = 0.52). Thus, the programs did improve several dimensions of ECE quality.

Children in schools assigned to the TT condition had improved outcomes in three of the four domains examined: early numeracy (d_{wt} =.11), early literacy (d_{wt} = .11), and social-emotional skills (d_{wt} = .18) in the first year. There were no impacts on executive function. There were no significant impacts of TTPA on children's outcomes. Compared to TT, children in TTPA had lower scores on domains of early literacy and numeracy, suggesting counteracting effects of the parental-awareness meetings. A second follow-up study (Wolf, Aber, Behrman, & Peele, 2019) found that a year later, in Year 2, impacts of TT were sustained only for social-emotional skills (d_{wt} = .13).

The Current Study

The current study builds the evidence-base of persistence and convergence of ECE programs by examining the issues for the first time in sub-Saharan Africa. The study extends previous findings on the QP4G programs in several ways. First, it provides evidence on long-term impacts of the interventions for one additional year when children were in Primary 1 (equivalent of first grade) and Primary 2 (equivalent of second grade) classes. It also considers whether the parental-awareness treatment of the QP4G programs, which was found to counteract the positive effects of the teacher-training on child outcomes, has persistent null effects. Second, outcomes cover domains of academic, social-emotional, and executive function skills, all foundational for longer-term school success, and multiple sources are used to measure children's outcomes. Third, this study examines contextual moderators of children's primary school classroom environments, including indicators of classroom process quality, extending evidence from U.S.-based research on the sustaining environments that support early intervention gains (Bailey et al., 2017). Fourth, this study extends conceptualizations of sustaining environments by examining teacher burnout as

a relevant indicator of subsequent educational contexts. The results have practical implications for the potential cascading effects of investing in pre-primary education in countries with limited resources and with extremely low primary education outcomes.

Method

The QP4G interventions were implemented between September 2015 and June 2016. Schools were randomly assigned to one of three treatment arms noted above: (a) TT, 82 schools; (b) TTPA, 79 schools; and (c) control, 79 schools. The school year in Ghana begins in September and ends in July. Data was collected in September–October 2015 (baseline), May–June 2016 (follow-up 1), May–June 2017 (follow-up 2), and May–June 2018 (follow-up 3). The present study used data collected at baseline and follow-up 3.

Randomization of schools was stratified by district and public and private sector. Six of the nine most disadvantaged districts (out of 16 districts total) in the Greater Accra Region were selected. The selected districts were rated as the most disadvantaged in the region based on the 2014 UNICEF District League Table (a social accountability index that ranks regions and districts based on development and delivery of key basic services, including education, health, sanitation, and governance; UNICEF, 2015). This trial was preregistered in the American Economic Association registry for randomized controlled trials (RCT ID: AEARCTR-0000704). 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").

Sampling and Data Collection Procedures

Schools were sampled from all six study districts in the summer of 2015. A listing of all schools was conducted using the Ghana Education Service Educational Management Information System database. Schools were then randomly sampled, stratified by district and within district by

public and private schools. Because there were fewer than 120 public schools across the six districts, every public school was sampled. Private schools (490 total) were sampled within districts in proportion to the total number of private schools in each district.

A subsample of children from each KG class was sampled within each school. Class rosters for all KG classrooms were collected, and an average of 15 children (eight from KG1, and seven from KG2) were randomly selected from each roster to participate in direct assessments. If a school had fewer than 15 children enrolled across both classrooms, all children were selected. For schools with only one KG classroom, 15 children were sampled. At baseline, the total sample of children was 3,435 children, with an average of 14.3 children per school (*range* = 4–15). Children (49.5% female) were, on average, 5.2 years old at baseline (*SD* = 1.2; For KG1, *M* = 4.8, *SD* = 1.1; and for KG2, *M* = 5.7, *SD* = 1.2). These children were followed at each subsequent wave of data collection. In Year 3 (*N* = 2,421), children were on average 7.8 years old.

In Year 3, the primary school teachers and classrooms of children were also sampled. Data were collected on teachers through administered surveys and classrooms through direct observation. These were new teachers who were not part of the original impact evaluation. Importantly, a large portion of the original sample (29.5%) could not be found at the third follow-up. Mobility and school dropout are common challenges to the school system in Ghana and many other LMICs (Kamanda & Sonkoh, 2015). The handling of missing data for this study is detailed below, and the results need to be interpreted in light of the large portion of missing data.

Measures

Child outcomes. Child outcomes were collected through direct assessments with trained multilingual assessors, teacher reports, and assessor reports. All assessors had experience working with children, received 5 days of training, and participated in 2 days of field practice. The

assessment tool was translated into three local languages (Twi, Ewe, and Ga). Details of the adaptation of assessments to the Ghanaian context are provided in the Supplementary Materials.

Academic skills. Literacy skills were measured across five domains of literacy and pre-literacy skills were measured primarily with the Early Grade Reading Assessment (RTI International, 2009a). Domains included expressive vocabulary, listening comprehension (in both English and the child's mother tongue), letter-sound identification, and nonword decoding. A measure of phonological awareness from the International Development and Early Learning Assessment (IDELA; Pisani, Dowd, & Borisova, 2018) was also included. The percent correct for each domain was computed, and the score for each domain averaged to create a total score ($M = 0.524$, $SD = 0.177$, $range = 0–0.932$, $\alpha = .76$).

Numeracy skills were measured across five domains of numeracy and pre-numeracy skills using the Early Grade Math Assessment (RTI International, 2009b). Domains included number identification, quantity discrimination, addition, subtraction, word problems, and missing number pattern identification. The percent correct for each domain was computed, and the score for each domain was averaged ($M = 0.470$, $SD = 0.172$, $range = 0–0.925$, $\alpha = .87$).

Social-emotional skills. Using the IDELA subscale (Pisani et al., 2018), prosocial skills were measured using 14 items grouped into five constructs: self-awareness, emotion identification, perspective taking and empathy, friendship, and conflict and problem solving. An example item of conflict solving involved asking children to imagine they are playing with a toy and that another child wants to play with the same toy, and then asking the children what they would do to resolve that conflict. "Correct" answers in the Ghanaian context, as agreed upon by the assessors during training, included talking to the child, taking turns, sharing, and getting another toy. The proportion correct for each domain was computed, and scores for each domain were averaged to create a total score ($M = 0.662$, $SD = 0.140$, $range = 0–1$, $\alpha = .67$).

Social competence was measured using The Social Competence Scale–Teacher Version (Conduct Problems Prevention Group, 1990), a 25-item measure that assesses a child's social behaviors (e.g., "cooperates with peers without prompting," "is helpful to others"), emotional self-regulation (e.g., "can accept things not going his or her way," "can wait patiently when necessary"), and academic-related behavioral skills (e.g., "functions well even with distractions," "stays on task"). The tool was pilot tested with Ghanaian teachers and found to be suitable for use with very minor adaptations (e.g., adding a brief explanation of some phrases). Items were scored on a scale where 1 = *not at all*, 2 = *a little*, 3 = *moderately well,* 4 = *well*, and 5 = *very well*. All items were averaged to create a total score and then re-scaled from 0 to 1 ($M = 0.692$, $SD = 0.138$, *range* = 0.20–1; $\alpha = .93$)

Executive function skills. *Cognitive flexibility* was assessed using an adapted version of the Dimensional Change Card Sort (DCCS)–Border version (Zelazo, 2006; 12 items, $\alpha = 0.78$). The DCCS is a standard procedure for assessing executive functioning early in development; in the more challenging border version used in this study, children need to follow different sets of sorting rules based on whether the card has a border. In the task, participants switch from sorting cards one way (e.g., by color) to sorting them a different way (e.g., by shape). *Inhibitory control* was measured using the Number Stroop Task (21 items, $\alpha = 0.96$). In this task, children are shown a set of boxes with one to four repeating numbers (e.g., 1111, 44), and are asked to report how many numbers are in each box; see Obradovíc et al. (2019) for details on the adapted measures). Scores within each domain were averaged to create a total percent correct score ($M = 0.545$, $SD = 0.173$, *range* = 0.00–1.00 for cognitive flexibility; $M = 0.930$, $SD = 0.139$, *range* = 0.27–1.00 for inhibitory control).

Finally, *behavioral regulation* was measured using an adapted version of the Preschool Self-Regulation Assessment–Assessor Report (PSRA-AS) (Smith-Donald et al., 2007). PSRA-AR

was designed to assess self-regulation of emotion, attention, and behavior of children during direct assessments. The instrument was shortened and adapted for the Zambian context (McCoy, Zuilkowski, & Fink, 2017) and used in other LMIC samples (Kim et al., 2018). The shortened version consisted of 13 items focused on the child's attention and behavior. Sample items include "pays attention during instructions and demonstrations," "remains in seat appropriately during test," and "modulates and regulates arousal level in self." Items were scored on a 1 to 4 scale and coded such that higher scores indicated better behavioral regulation. Total scores were re-scaled to 0 to 1 ($M = 0.779$, $SD = 0.062$, $range = .429–.944$; $\alpha = .88$).

Classroom quality. Children's primary school classrooms ($N = 841$ out of 911 total, or 92.3%) were videotaped during a lesson for 30–60 minutes. Videos were coded using the Teacher Instructional Practices and Processes System-Primary School Version (TIPPS; Seidman, Kim, Raza, Ishihara & Halpin, 2018), with minor adaptations for use in Ghana (e.g., referring to pupils as children, as is common in Ghanaian kindergarten settings). The TIPPS is a classroom observation tool for assessing classroom quality that focuses on the nature of teacher–child interactions; it was created for use in LMICs. The TIPPS is made up of 19 items and was validated for the Ghanaian context in a previous study (Wolf et al., 2018). All raters were Ghanaian nationals and had at least a bachelor's degree. After completing a training, each rater had to meet or exceed calibration criteria within three attempts to be certified as a TIPPS observer (see Wolf et al. (2018) for further details). To assess inter-rater reliability, 15% of videos were coded by two additional raters (three total). The intraclass correlation coefficient (ICC) of the final scores was calculated to assess how the partition of variance in scores broke down into differences in individual raters and how much variance was shared across raters. On average across items, 70.9% of the variance was shared across raters, denoting acceptable inter-rater reliability.

A factor analysis was conducted to examine underlying dimensions of quality. The dataset

was randomly split into an exploratory ($N = 419$) and confirmatory ($N = 422$) sample. The purpose of the exploratory sample was to allow for multiple variations of the 18 items on initial models to be fitted in order to arrive at a proposed model. The purpose of the confirmatory sample was to ensure that the proposed model demonstrated out-of-sample generalizability. Analyses were conducted using MPlus V6.12. One item was dropped due to lack of variation (i.e., "teacher shows favoritism to some students over others"). Factor solutions were examined for the remaining 17 items ranging from one to four factors in the exploratory analysis using varimax rotation. In selecting a model, I considered (a) goodness-of-fit statistics and (b) conceptual meaning of the factors. The selected model from the exploratory analysis was a two-factor model, which had adequate goodness-of-fit statistics in the confirmatory sample ($\chi^2(89) = 256.03$, CFI $= .94$, TLI $= .93$, RMSEA $= .067$).

The two factors were *instructional support* (8 items; e.g., students' ideas and interests considered in lesson; shows relevance of lesson; specific quality feedback; $\alpha = .76$) and *emotional support* (7 items; e.g., positive environment, tone of voice; effective behavior management; $\alpha = .81$). The scores for each scale were computed through an average of the items and were normally distributed ($M = 1.96$, $SD = 0.44$, *range* $= 1$–3.50, $\alpha = .76$; and $M = 3.34$, $SD = 0.46$, *range* $= 1.71$–4, $\alpha = .81$, respectively) and positively correlated ($r = 0.16$). The intraclass correlations (ICC) of instructional support and emotional support were 0.296 and 0.131, respectively.

Teacher burnout. Teachers of children's primary classrooms ($N = 879$ out of 911 total, or 96.5%) were administered a survey in English (the language of Ghana's education system) and answered items from the Maslach Burnout Inventory (MBI; Maslach et al., 1996). The nine items were from the *Emotional Exhaustion* subscale and were normally distributed. Teachers answered each item on a scale from 0 ("never") to 6 ("every day") to indicate, for instance, how often they have felt "Frustrated," "Burned out," "Stressed," and "Emotionally drained" ($M = 1.65$, $SD = 1.14$,

range = 0–5.8, α = .78). A previous study assessing the factor structure of the MBI with teachers in Ghana has shown that the items hold as a single factor, and that this factor has high concurrent validity with teachers' job demands and job resources (Wolf & Lee, 2019). The ICC for burnout was 0.128. It was not correlated with either dimension of classroom quality (r = .04 with instructional support, r = -.01 with emotional support).

Covariates. Covariates were included to improve the precision of the estimates and were drawn from the baseline data: private-sector status of the school, six district dummies, child sex, baseline age in years, and KG level at baseline (1 = KG1, 2 = KG2, or 3 = combined classroom). Baseline scores, measured using the IDELA (Pisani et al., 2018), were available for all domains (i.e., literacy, numeracy, social-emotional, and executive function). Scores were standardized and included in the respective models.

Analytic Plan

Baseline equivalence and attrition analysis. A baseline equivalency analysis was conducted in an earlier study to confirm whether the randomization was successful in yielding statistically equivalent treatment and control groups (Wolf et al., 2019). Given attrition in the sample over the 3 years of the study, I conducted a differential attrition analysis for the three treatment groups. Nearly 30% (1,014) of children from baseline were not in the follow-up 3 sample (34.5% of control; 29.3% of TT, and 30.5% of TTPA). Treatment status significantly predicted attrition; children in both treatment groups were more likely to be in the sample compared to the control group (TT, $\chi^2(1)$ = 7.92, p < .01; TTPA, $\chi^2(1)$ = 4.80, p < .05).

Standards developed by the What Works Clearinghouse (WWC), established by the U.S. Department of Education's Institute of Education Sciences, classify evaluation studies as having either "high" or "low" attrition based on a combination of overall and differential attrition. The WWC standards also account for an important trade-off between overall and differential attrition—

namely, that a study can have a higher overall rate of attrition if it has a low rate of differential attrition. In this case, the combination of overall attrition at 29.5%, and differential attrition of 5.2% in TT, and 4.1% in TTPA, categorize this in the range of low attrition using the conservative attrition standard (WWC, 2017). Some data were also missing on classroom contexts (7.7%) and teacher professional well-being (3.5%) due to teacher absence or refusal.

Missing data imputation. Following the WWC Version 4.0 Standards Handbook (IES, 2017), I used multiple imputation to address missing data for all missing cases, including dependent variables and treatment status, using four rounds of data collection (baseline and all three follow-ups). While the data were not missing completely at random (MCAR), if variables that strongly predict attrition are incorporated into the missing data strategy, the plausibility of the missing-at-random (MAR) assumption increases (Young & Johnson, 2015). In other words, by using a large set of covariates in estimating multiple chains of models, including those that predict differential attrition, assumptions of MAR have been shown to be robust. Impact estimates were computed using 100 imputed datasets.[2]

Impact analysis. Multiple observations arising from the same sampling unit (in this case, schools and classrooms) will have model residuals that are correlated. Multilevel models (also known as random effects models) provide flexible and advantageous strategies for modeling such dependencies. To account for the nested, non-independent nature of the data when children were sampled at baseline from schools and classrooms, I employed three-level models to estimate impacts on child-development outcomes. Impact analyses included a select set of covariates.

---

[2] Imputation was conducted in two steps. First, 20 teacher-level datasets were imputed using a rich set of teacher demographic and background variables, outcome scores for professional well-being and classroom quality across all four waves, and treatment status indicators. Second, 10 of these teacher datasets were randomly sampled. Each individual dataset was merged with the child outcomes data, caregiver data, and basic children demographic characteristics from all waves of data. For each of the 10 teacher datasets, 10 child datasets were imputed, resulting in 100 child-level datasets.

Children were nested in the classrooms and schools from which they were sampled. The multiply-imputed datasets were analyzed using Rubin's combining rules to compute pooled coefficients and standard errors across datasets (Rubin, 1987). I estimated treatment impacts on each outcome using the following equation:

Level 1 (Child-level) Model:

$$Y_{ijk} = B_{0jk} + B_{1jk}'X_{ijk} + e_{ijk}$$

Where $\mathbf{X_{ijk}}$ is the vector of child covariates (gender, age, and baseline score).

Level 2 (Classroom-level) Model:

$$B_{0jk} = \gamma 00k + u0jk$$

Where $\mathbf{B_{0jk}}$ is the classroom-level random intercept.

Level 3 (School-level) Model:

$$\gamma_{00k} = \pi_{000} + \pi_{001}TT_k + \pi_{002}TTPA_k + \pi_{003}'Z_k + v_{00k}$$

Where $\mathbf{\gamma_{00k}}$ is the school-level random intercept; $\mathbf{Z_k}$ is the vector of school-level covariates (district dummies, private or public, and four dummy variables for different school-mobility scenarios); and $\mathbf{TT_k}$ and $\mathbf{TTPA_k}$ are indicators for schools assigned to the two treatment arms.

Next, I examined whether sustained intervention impacts were moderated by three dimensions of children's primary classroom context by adding interaction terms between each treatment condition and the primary classroom variable. On average, there were 2.7 children per primary school classroom (*range* = 1–16). Given the sampling strategy used in the initial project and in previous studies (Wolf, Aber, Behrman, & Tsinigo, 2019; Wolf, Aber, Behrman, & Peele, 2019), for the analysis I nested children in the pre-primary schools and classrooms from which they were sampled. Thus, the primary school classroom / teacher variables are modeled at the child level.

Results

The ICCs for children's primary school outcomes based on children's pre-primary schools and classrooms are displayed Table 2. For academic outcomes, a larger proportion of variance is attributed to the school and classroom-levels than for the nonacademic outcomes. For literacy, for example, 23.5% of the variance is explained by the pre-primary school level, 12.2% by the pre-primary grade level, and 64.4% by the child level. In contrast, for social-emotional outcomes, 8.8% of the variance is explained by the pre-primary school level, 6.0% by the pre-primary grade level, and 85.2% by the child level.

Persistent Impacts of QP4G on Child Development

Table 3 presents impact estimates of QP4G on children's developmental outcomes 2 years after the end of the school-year-long program, with each row representing the average effect for each treatment condition relative to the control group. For academic outcomes, there were marginally statistically significant persistent impacts of TT on children's literacy ($b = 0.019$, $SE = 0.010$, $p < 0.06$, $d_{wt}^{3} = .11$). There were statistically significant persistent negative impacts of TTPA on numeracy skills ($b = -0.016$, $SE = 0.008$, $p < 0.05$, $d_{wt} = -.10$). For social-emotional outcomes, impact estimates for prosocial and social competence skills did not reach traditional levels of statistical significance. There were marginally significant impacts of TT on social competence ($b = 0.015$, $SE = 0.009$, $p < 0.10$, $d_{wt} = .11$). For executive function skills, there were statistically significant impacts of TT on cognitive flexibility ($b = 0.022$, $SE = 0.009$, $p < 0.05$, $d_{wt} = .12$), and

---

[3] d_{WT} represents a standardized mean difference between treatment and control schools. This was calculated with the following equation from Hedges (2009):

$$d_{WT} = \frac{b}{\sqrt{\hat{\sigma}_{BS}^2 + \hat{\sigma}_{BC}^2 + \hat{\sigma}_{WC}^2}},$$

where b represents the unstandardized regression coefficient with covariate adjustment (e.g., $b = .018$), and the three terms of the denominator represent variances at the school, classroom/teacher, and child levels, respectively, without covariate adjustment. The rationale behind covariate adjustment for the treatment effect, but not the variances, was to obtain a more precise treatment effect (i.e., adjusted), but standardized based on typical (i.e., unadjusted) variances at each level (L. V. Hedges, personal communication, November 3, 2014). Variance estimates for each level were computed using the pooled estimates across all imputed datasets using Rubin's combining rules.

marginally statistically significant impacts of TT on inhibitory control ($b = 0.015$, $SE = 0.009$, $p < 0.10$, $d_{wt} = .08$). There were also statistically significant impacts on children's behavioral regulation for both TT ($b = 0.015$, $SE = 0.006$, $p < 0.01$, $d_{wt} = .15$) and TTPA ($b = 0.014$, $SE = 0.006$, $p < 0.05$, $d_{wt} = .14$).

To test differences between the TT and TTPA conditions, all models were re-run with TTPA as the reference group. There were significant differences between the two treatments, including in impacts on literacy ($b = 0.026$, $SE = 0.010$, $p < 0.01$) and numeracy ($b = 0.016$, $SE = .007$, $p < .05$). There were no differences in social-emotional or executive function outcomes.

Moderation of Persistent Impacts by Classroom Environment

For classroom environment, I examined instructional support, emotional support, and teacher burnout. The results are presented in Table 4. There were no moderating effects of classroom instructional support on any outcome. There was evidence of moderating effects of classroom emotional support on numeracy outcomes for both TT ($b = 0.028$, $SE = 0.015$, $p < .07$) and TTPA ($b = 0.030$, $SE = 0.015$, $p < .05$). The results are graphed in Figure 1 for ease of interpretation, and indicate that for TT, impacts of the treatment were positive and largest in classrooms with relatively high levels of emotional support. For TTPA, a similar pattern is observed, though the figure also shows that the persistent negative impacts on numeracy were concentrated in classrooms with low levels of emotional support.

There was also evidence that teacher burnout moderated persistent impacts on academic outcomes. This was the case for literacy outcomes for both TT ($b = -0.020$, $SE = 0.007$, $p < .01$) and TTPA ($b = -0.013$, $SE = 0.007$, $p < .06$), and for numeracy outcomes for TT ($b = -0.013$, $SE = 0.006$, $p < .05$). The results are graphed in Figures 2 and 3 and indicate that there were positive persistent impacts on both outcomes in classrooms where teachers had lower burnout levels. There were no moderating effects for social-emotional or executive function outcomes.

Discussion

This study addressed two critical questions about the influences of early childhood education on child development that have not been examined before in SSA, building on two previous studies from a school-randomized trial of a preschool teacher-training program in Ghana. In the previous studies, the training of preschool teachers was found to have positive impacts on children's literacy, numeracy, and social-emotional skills in Year 1 during implementation (Wolf, Aber, Behrman, & Tsinigo, 2019); and only impacts on children's social-emotional skills persisting in Year 2 (Wolf, Aber, Behrman, & Peele, 2019). When the same training was implemented with a parental-awareness intervention, there were no positive impacts on children.

This present study extended these results to examine, first, if any impacts on children's developmental skills were sustained in Year 3 in three key developmental domains: academics (literacy and numeracy), social-emotional skills (i.e., prosocial and social competence), and executive function (i.e., cognitive flexibility, inhibitory control, and behavioral regulation). As such, this study constitutes the first test of persistence or convergence of a preschool teacher professional development program on children's primary school outcomes in an SSA country. This study also examined if there were different patterns of sustained impacts based on the quality of children's primary school classrooms. This is the first study to assess the role of "sustaining environments" in persistence of ECE impacts in the SSA region. Two important strengths are: (a) the measurement of quality using both observational data on process quality (as opposed to structural indicators), and (2) expanding the conceptual and operational definition of quality to include teacher burnout. To address these questions, I followed a large longitudinal sample of students originally in the two grades of pre-primary school in the Greater Accra region of Ghana at the end of their Primary 1 or Primary 2 school year.

Persistent Impacts on Academic Outcomes

There was weak evidence for persistence of impacts on children's academic outcomes for the full sample. In the Year 3 follow-up, there were marginally statistically significant positive impacts of the teacher-training treatment (TT) on literacy skills ($d_{wt} = 0.11$, $p < .06$), and a small but significant *negative* impact ($d_{wt} = -0.10$, $p < .05$) of the same treatment delivered with the parental-awareness (TTPA) intervention. The impacts on academic outcomes in Year 3 may be due to the increased academic demands of primary school, as children adjust their behavior to changing demands from their environment (Lourenco & Casey, 2013), and these changes may have been affected by earlier skill differences between the treatment groups. Alternatively, the appearance of impacts in Year 3 might be due to measurement-related issues. In the Year 1 and 2 follow-up studies, the same assessment battery was used to measure literacy and numeracy outcomes. In the Year 3 follow-up, however, a new assessment battery of literacy and numeracy was used given children's growth and developmental changes. It is possible that the differential significance of impact findings across years is in fact due to the different measures. Given the relatively small impacts found in Year 1 of the program ($d_{wt} = .11$ in both literacy and numeracy), the small magnitude of these persistence rates is not surprising. Importantly, there were significant differences in both outcomes contrasting TT and TTPA, with children in TT outperforming those in TTPA in literacy ($d_{wt} = .15$, p $< .01$) and numeracy ($d_{wt} = .09$, $p < .05$).

The negative impacts of TTPA on numeracy are concerning, as they suggest that whatever effect the treatment had on parents during the implementation of the program may have led to sustained changes in parenting practices that resulted in worse outcomes for children. Importantly, the magnitude of the impact is small. But interpreting these effects requires a deeper discussion of the context in which the program was implemented. Previous studies have shown that Ghanaian parents' value early education and demand academically focused, rigorous instruction from teachers (Bidwell et al., 2014; Kabay et al., 2017). A recent study on parents' school involvement

has been shown to be negatively predictive of Ghanaian children's school-readiness skills (Wolf & McCoy, 2019), suggesting that parents may have a vision for schooling that is in contradiction to developmental learning processes. Another study in Tanzania found that parents consider respect and social compliance as core values they hope schooling will instill in their children, while teachers value confidence and curiosity (Jukes et al., 2018). It is possible that parents compensated for the changes in their child's preschool by becoming more actively involved at home, ultimately to the detriment of their child's learning.

There may be other explanations that are related to the study design. For example, given the large sample size and the small magnitude of the effect, it is possible that the negative impact detected in this study was a Type I error. Alternatively, while baseline equivalence was established (Wolf, Aber, Behrman & Tsinigo, 2019), it possible that the treatment groups were not balanced on variables that were not measured or observed in ways that could have created the impact independent of the treatment. Given that these findings are quite unusual, these explanations are worth considering as well. Nonetheless, more research is needed to understand Ghanaian parents' desires and goals for their child's socialization and education, and the ways that they attempt to support these processes, in order to ensure that future interventions builds on these desires in positive ways.

Persistent Impacts on Social-Emotional and Executive Function Outcomes

There was evidence of sustained impacts on social-emotional and executive function outcomes. There were small positive impacts (d_{wt} = .08-.15) on all outcomes examined in these domains, though with variable precision, with four of the five outcomes reaching statistical significance at $p < .10$. Impacts on prosocial skills, which were measured through direct assessments with trained data collectors, were not statistically significant. Impacts on teacher-reported social competence were marginally statistically significant in the TT treatment condition

($d_{wt} = 0.11$, $p < .10$). Given the relatively small impacts on social-emotional skills in Year 1 of the program ($d_{wt} = .18$) and Year 2 ($d_{wt} = .13$), the magnitude of the persistence rates is relatively consistent.

Furthermore, there were sustained impacts of the TT condition on executive function skills, specifically cognitive flexibility ($d_{wt} = 0.12$, $p < .05$), and marginally significant impacts on inhibitory control ($d_{wt} = 0.08$, $p < .08$). Finally, there were impacts of both treatment conditions on assessor-reported behavioral regulation ($d_{wt} = 0.15$ and 0.14 for TT and TTPA, respectively). It is promising that there is some evidence for sustained (albeit small) impacts on executive function skills, as some have argued that this is a key channel for longer-term impacts of ECE interventions (Raver, 2003; Raver et al., 2011). The pattern of results differs from recent studies in the United States. Several such studies found no detectable effects of preschool shortly after kindergarten (Hill et al., 2015; Puma et al., 2012). Interestingly, a few recent studies found that children who attended preschool outperformed their peers academically but had *lower* psychosocial skills in primary school (e.g., Ansari, 2018), or that positive impacts on academic and behavioral outcomes reversed in early primary school (Lipsey, Farran, & Durkin, 2018).

Sustaining Environments: Variation by Primary School Classroom Environments

I also examined variation in persistent effects based on children's primary classroom environments. Studies testing sustaining environments have used a range of measures to define quality, including both structural and process features of classrooms and schools. In this study, sustaining environments were measured through direct observations of teacher–child interactions during classroom instruction (instructional and emotional support) and teacher burnout levels. These three elements of the classroom context have not yet been examined directly in the literature on sustaining environments, though evidence suggests they are key drivers for child development and learning (e.g., Hamre et al., 2013; Pianta et al., 2005).

There was supporting evidence that higher quality environments were more likely to yield persistent gains from exposure to quality ECE for academic outcomes. Despite no main effects of TT on numeracy, the pattern of findings showed small positive, persistent impacts for children in classrooms with high levels of emotional support. Similarly, the negative impacts of TTPA on numeracy were concentrated for children in classrooms with low emotional support. It is possible that there was no convergence in academic skills in higher quality classroom environments because better teachers build on children's existing skills. Because teachers in Ghana tend to focus on academic skill development (Bidwell et al., 2014), this may have further ingrained differences that may have existed at the start of the school year.

There were no differences in sustained impacts based on classroom instructional support, suggesting that in the context of this intervention, continued classroom emotional support (which also improved during the intervention year in children's ECE classrooms) was more consequential in buttressing previous gains from the intervention. The results support and extend the findings that alignment between ECE and early primary school classroom environments may be an important factor in sustaining impacts from ECE (Clements et al., 2013). Furthermore, the results suggest that cumulative high-quality classroom environments in the preschool and early elementary school years can lead to better developmental outcomes for children (Carr, Mokrova, Vernon-Feagans, & Burchinal, 2019), and that cumulative poor-quality environments in this period can be detrimental.

Finally, I examined teacher burnout as an additional consequential indicator of the classroom context given its relevance to LMICs (e.g., Bennell & Akyeampong, 2007; Wolf & Lee, 2019). The original QP4G teacher-training reduced teacher burnout and improved nurturing teacher–child interactions. This element of the classroom environment may be of particular relevance when considering the contrast between children's ECE and primary classroom contexts.

Impacts on both literacy and numeracy skills were positive in classrooms where teachers had low levels of burnout, and negative impacts of TTPA on numeracy were larger in classrooms with teachers with high levels of burnout. These findings build support for examining teacher burnout more broadly in education research, including a recent study finding that burnout was directly associated with lower student achievement (McLean & Connor, 2015). Teacher burnout was not correlated with the teacher–child interactions observed during the classroom observation period, suggesting that burnout may affect interactions outside of formal instructional time in ways that may hinder children's development and learning. The results highlight an important area for research on sustaining environments to consider teacher burnout and the potential pathways through which it may moderate sustained preschool impacts. Future research may also consider the joint effects of teacher burnout and classroom quality to more holistically depict children's subsequent schooling experiences. Given that the constructs were uncorrelated in this dataset, understanding the potential cumulative consequences of both would be important.

Interestingly, variation in persistent impacts was only found for academic skills, while the sustained impacts on nonacademic skills did not depend on subsequent school environments. Bailey, Duncan, Odgers, and Yu (2017) argued that interventions that target (directly or indirectly) skills in children that are malleable, fundamental, and would not develop over time in the absence of the intervention are more likely to yield persistent impacts. They characterize literacy and numeracy skills, as well as social-emotional and executive function skills, as both malleable and fundamental. But they claim rudimentary literacy and numeracy skills are among those that grow rapidly in most counterfactual conditions. In contrast, they identify other skills that may not "fade-out," especially in more adverse environments, including some normative cognitive functions such as executive functions. Bailey and colleagues argued that skills that meet all three criteria (malleable, fundamental, would not develop normally in the absence of intervention) are more

likely to persist. Using this framework, perhaps the executive function skills measured in this study met all three criteria and represent these "trifecta skills" in the Ghanaian school context.

Limitations and Conclusions

The findings of this study must be interpreted in light of its limitations. First, without further exploration, the results cannot be generalized beyond the six districts in the Greater Accra Region in Ghana included in the sample, nor outside of Ghana, a country with a unique ECE context in the region. Second, there was significant attrition of the children in the sample (about 30% of the baseline sample). The use of multiple imputation and multiple controls probably limits bias due to attrition. Notably, the magnitude and pattern of results are similar for the sample of children who were present both at baseline and in Year 3, but the pattern of statistical significance is weaker (see Table S1). Third, due to resource constraints, very few data were collected on the implementation of the parental-awareness meetings and parents' engagement in and perceptions of this training. Thus, the unexpected findings of the parental-awareness meetings are difficult to explain and limited to speculation.

Nonetheless, the study advances the field of global ECE and child development. Global initiatives such as the Sustainable Development Goals aim to reduce inequalities in child development by expanding access to learning opportunities, with a specific target for early childhood education (United Nations, 2015). At the same time, little evidence exists on how quality early learning programs impact children's outcomes in the longer term in SSA. The findings suggest that a brief teacher training, built into existing governmental systems and implemented over the course of 1 school year, can lead to longer-term gains for children, particularly in domains that may not otherwise be promoted in the Ghanaian schooling system. Importantly, the small and relatively weak results for the full sample suggest that improvements to the training program are necessary to truly address the educational challenges Ghana faces.

While the effect sizes of QP4G are considered small by standards of experimental lab-based research, some have argued that these standards are not relevant for school-based educational interventions. Kraft (2018) proposed new standards for interpreting effect sizes, with $<.05$ = small, $.05$–$.20$ = medium, and $>.2$ = large. Furthermore, a focus on magnitude alone obscures important differences in study features, program costs, and scalability in education research. Effect sizes should also be considered with program cost and scalability. It is worth noting that there were impacts of QP4G 2 years after the programs ended. This is notable for a program that was designed with national scalability in mind, thus limiting its intensity and cost. Cost analysis of the QP4G program showed that the teacher-training and coaching component cost 21.17 USD per child (or 529.17 USD per teacher, assuming 25 students per teacher). The tradeoff between achieving large impacts and creating interventions that are affordable and can feasibly be implemented at scale must be considered when interpreting short- and long-term impacts of educational interventions.

The findings from this study support the hypothesis that longer-term impacts of quality ECE depend, to some extent, on the quality of children's future schooling environments, and point to an expanded set on indicators to represent important elements of a sustaining classroom context. Governments in low-income countries are expanding investments in pre-primary education and in improving primary education, but these efforts are rarely aligned. The results suggest that if these investments can be coordinated, there is better potential for scalable, long-term gains in children's learning and development. Close to 90% of children in the world live in a low- or middle-income country (World Bank, 2016), and these children are attending preschool at growing rates (McCoy et al., 2018). More studies examining children's experiences in ECE and transition to primary schooling in developing countries are needed if global efforts to improve educational quality and child development around the world are to succeed.

References

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

Ansari, A. (2018). The persistence of preschool effects from early childhood through adolescence. *Journal of Educational Psychology, 110*(7), 952–973. doi:10.1037/edu0000255

Ansari, A., & Pianta, R. C. (2018). The role of elementary school quality in the persistence of preschool effects. *Children and Youth Services Review, 86*, 120–127. doi:10.1016/j.childyouth.2018.01.025

Araujo, M. C., Carneiro, P., Cruz-Aguayo, Y., & Schady, N. (2016). Teacher quality and learning outcomes in kindergarten. *Quarterly Journal of Economics, 131*(3), 1415–1453. doi:10.1093/qje/qjw016

Arnold, D. H., Kupersmidt, J. B., Voegler-Lee, M. E., & Marshall, N. A. (2012). The association between preschool children's social functioning and their emergent academic skills. *Early Childhood Research Quarterly*, *27*(3), 376–386. doi:10.1016/j.ecresq.2011.12.009

Bailey, D., Duncan, G. J., Odgers, C. L., & Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research on Educational Effectiveness,10*, 7–39. doi:10.1080/19345747.2016.1232459

Bassok, D., Gibbs, C. R., & Latham, S. (2015). *Do the effects of early childhood interventions systematically fade? Exploring variation in the persistence of preschool effects*. Retrieved from http://curry.virginia.edu/uploads/resourceLibrary/36_Preschool_Fade_Out.pdf

Bennell, P., & Akyeampong, K. (2007). *Teacher motivation in Sub Saharan Africa and South Asia* (Educational Papers no. 71). Department for International Development.

Berlinski, S., Galiani, S., & Gertler, P. (2009). The effect of pre-primary education on primary school performance. *Journal of Public Economics, 93*(1), 219–234. doi:10.1016/j.jpubeco.2008.09.002

Bidwell, K., Parry, K., & Watine, L. (2014). *Exploring Early Education Programs in Peri-urban Settings in Africa: Accra, Ghana.* UBS Optimus Foundation.

Black, M. M., Walker, S. P., Fernald, L. C., Andersen, C. T., Digirolamo, A. M., Lu, C., . . . Grantham-

Mcgregor, S. (2017). Early childhood development coming of age: Science through the life course. *The Lancet, 389*(10064), 77–90. doi:10.1016/S0140-6736(16)31389-7

Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *American Psychologist*, 57(2), 111–127.

Blair, C., & Raver, C.C. (2014). Closing the achievement gap through modification of neurocognitive and neuroendocrine function: Results from a cluster randomized controlled trial of an innovative approach to the education of children in kindergarten. *PLoS ONE,* e112393. doi:10.1371/journal.pone.0112393

Bouguen, A., Huang, Y., Kremer, M. R., & Miguel, E. (2019). *Using RCTs to estimate long-run impacts in development economics* (CEPR Discussion Paper No. DP13443).

Bull, R., & Lee, K. (2014). Executive functioning and mathematics achievement. *Child Development Perspectives, 8(1),* 36–41. doi:10.1111/cdep.12059

Campbell, F. A., Pungello, E. P., Burchinal, M., Kainz, K., Pan, Y., Wasik, B. H., . . . Ramey, C. T. (2012). Adult outcomes as a function of an early childhood educational program: An Abecedarian Project follow-up. *Developmental Psychology,48*(4), 1033–1043. doi:10.1037/a0026644

Carr, R. C., Mokrova, I. L., Vernon-Feagans, L., & Burchinal, M. R. (2019). Cumulative classroom quality during pre-kindergarten and kindergarten and children's language, literacy, and mathematics skills. *Early Childhood Research Quarterly,47*, 218–228. doi:10.1016/j.ecresq.2018.12.010

CASEL. (2017). *Core SEL Competencies.* Retrieved from http://www.casel.org/core-competencies/

Clements, D. H., Sarama, J., & Germeroth, C. (2016). Learning executive function and early mathematics: Directions of causal relations. *Early Childhood Research Quarterly*, *36*, 79–90. doi:10.1016/j.ecresq.2015.12.009

Clements, D. H., Sarama, J., Wolfe, C. B., & Spitler, M. E. (2013). Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies: Persistence of effects in the third year. *American Educational Research Journal, 50*(4), 812–850. doi:10.3102/0002831212469270

Conduct Problems Prevention Research Group. (1990). *Social Competence Scale (teacher version).* Fast

Track Project. Retrieved from http://www.fasttrackproject.org

Curby, T. W., Brown, C. A., Bassett, H. H., & Denham, S. A. (2015). Associations between preschoolers' social–emotional competence and preliteracy skills. *Infant & Child Development, 24(5),* 549–570. doi:10.1002/icd.1899

Currie, J., & Thomas, D. (2000). School quality and the longer-term effect of head start. *Journal of Human Resources, 35*(4), 755–774. doi:10.2307/146372

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science, 318*(5855) 1387–1388. doi:10.1126/science.ll511.

Dobbs, J., Doctoroff, G. L., Fisher, P. H., & Arnold, D. H. (2006). The association between preschool children's socio-emotional functioning and their mathematical skills. *Journal of Applied Developmental Psychology*, *27*(2), 97-108. doi: 10.1016/j.appdev.2005.12.008

Dodge, Bai, Ladd, & Muschkin, (2017). Impact of North Carolina's early childhood programs and policies on educational outcomes in elementary school. *Child Development, 88(3),* 996-1014. doi: 10.1111/cdev.12645.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ... & Sexton, H. (2007). School readiness and later achievement. *Developmental Psychology*, *43(6)*, 1428–1446. doi:10.1037/0012-1649.43.6.1428

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

Engle, P. L., Fernald, L. C., Alderman, H., Behrman, J., Ogara, C.,… Iltus, S. (2011). Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries. *The Lancet, 378*(9799), 1339–1353. doi:10.1016/S0140-6736(11)60889-1

Ganimian, A. J., & Murnane, R. J. (2016). Improving education in developing countries: Lessons from rigorous impact evaluations. *Review of Educational Research, 86*(3), 719–755. doi:10.3102/0034654315627499

Ghana Education Service. (2012). *Programme to scale-up quality kindergarten education in Ghana.* Ghana

    Ministry of Education. Accra, Ghana. Retrieved from https://issuu.com/sabre-

    tom/docs/10_12_12_final_version_of_narrative_op

Ghana Ministry of Education. (2014). *Ghana 2013 national education assessment: Technical report*.

Ghana Ministry of Education. (2016). *Education sector performance report, 2016*. Accra, Ghana.

Gormley, W. T., Phillips, D., & Anderson, S. (2017). The effects of Tulsa's pre-K program on middle

    school student performance. *Journal of Policy Analysis and Management, 37*(1), 63–87.

    doi:10.1002/pam.22023

Hamre, B. K., Pianta, R. C., Downer, J. T., DeCoster, J., Mashburn, A. J., Jones, S. M., . . .& Brackett, M.

    A. (2013). Teaching through interactions. *The Elementary School Journal, 113(4),* 461–487.

    doi:10.1086/669616

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P., & Yavitz, A. (2010). The rate of return to the

    High/Scope Perry Preschool program. *Journal of Public Economics, 94*(1), 114–128.

    doi:10.1016/j.jpubeco.2009.11.001

Hill, C. J., Gormley, W. T., & Adelstein, S. (2015). Do the short-term effects of a high-quality preschool

    program persist? *Early Childhood Research Quarterly, 32*, 60–79. doi:10.1016/j.ecresq.2014.12.005

Howes, C., James, J., & Ritchie, S. (2003). Pathways to effective teaching. *Early Childhood Research*

    Quarterly, 18, 104–120. doi:10.1016/S0885-2006(03)00008-5

Jenkins, J. M., Watts, T. W., Magnuson, K., Gershoff, E. T., Clements, D. H., Sarama, J., & Duncan, G. J.

    (2018). Do high-quality kindergarten and first-grade classrooms mitigate preschool fadeout? *Journal of*

    Research on Educational Effectiveness, 11, 339–374. doi:10.1080/19345747.2018.1441347

Johnson, R. (2013). *School quality and the long-run effects of Head Start*. Working paper. Retrieved from

    http://www.nber.org.proxy.library.upenn.edu/papers/w156664

Jukes, M., Gabrieli, P., Mgonda, N. L., Nsolezi, F., Jeremiah, G., Tibenda, J., & Bub, K. L. (2018).

    "Respect is an Investment": Community Perceptions of Social and Emotional Competencies in Early

Childhood from Mtwara, Tanzania. *Global Education Review*, 5(2), 160-188.

Kabay, S., Wolf, S., & Yoshikawa, H. (2017). "So that his mind will open": Parental perceptions of preschool in urbanizing Ghana. *International Journal of Educational Development, 57* ,44-53. https://doi.org/10.1016/j.ijedudev.2017.08.013

Kamanda, M., & Sonkoh, O. (2015). A systematic review of the status of children's school access in low- and middle-income countries between 1998 and 2013: Using the INDEPTH Network platform to fill the research gaps. *Global Health Action, 8,* e28430. doi:10.3402/gha.v8.28430

Kim, H.Y., Brown, L., Tubbs Dolan, C., Sheridan, M., & Aber, J.L. (2018, under review). PostMigration Risk Factors, Developmental Processes, and Learning Outcomes among Syrian Refugee Children.

Kraft, M. A. (2018). *Interpreting effect sizes of education interventions.* Working paper. https://scholar.harvard.edu/files/mkraft/files/kraft_2018_interpreting_effect_sizes.pdf

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research, 88*(4), 547–588. doi:10.3102/0034654318759268

Lee, V. E., & Loeb, S. (1995). Where do Head Start attendees end up? One reason why preschool effects fade out. *Educational Evaluation and Policy Analysis, 17*(1), 62–82. doi:10.3102/01623737017001062

Lee, S., & Wolf, S. (2019). Measuring and predicting burnout among early childhood educators in Ghana. *Teaching and Teacher Education, 78,* 49-61. https://doi.org/10.1016/j.tate.2018.10.021

Lipsey, M. W., Farran, D. C., & Durkin, K. (2018). Effects of the Tennessee Prekindergarten Program on children's achievement and behavior through third grade. *Early Childhood Research Quarterly, 45*(4)*,* 155–176. doi:10.1016/j.ecresq.2018.03.005

Magnuson, K. A., Ruhm, C., & Waldfogel, J. (2007). The persistence of preschool effects: Do subsequent classroom experiences matter? *Early Childhood Research Quarterly, 22*(1):18–38. doi:10.1016/j.ecresq.2006.10.002.

Martinez, S., Naudeu, S., & Pereira, V. (2012). *The promise of preschool in Africa: A randomised impact evaluation of early childhood development in rural Mozambique.* World Bank Group.

Maslach, C., Jackson, S. E., & Leiter, M. P. (1996). *Maslach Burnout Inventory Manual* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.

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. *PLoS Medicine, 13*(6), e1002034. doi:10.1371/journal.pmed.1002034

McCoy, D. C., Salhi, C., Yoshikawa, H., Black, M., Britto, P., & Fink, G. (2018). Home- and center-based learning opportunities for preschoolers in low- and middle-income countries. *Child and Youth Services Review, 88,* 44–56. doi:10.1016/j.childyouth.2018.02.021

McCoy, D. C., Yoshikawa, H., Ziol-Guest, K. M., Duncan, G. J., Schindler, H. S., …Shonkoff, J. P. (2017). Impacts of early childhood education on medium- and long-term educational outcomes. *Educational Researcher, 46,* 474–487. doi:10.3102/0013189X17737739

McCoy, D. C., Zuilkowski, S. S., & Fink, G. (2015). Poverty, physical stature, and cognitive skills: Mechanisms underlying children's school enrollment in Zambia. *Developmental Psychology*, *51*(5), 600–614. doi:10.1037/a0038924

McLean, L., & Connor, C. M. (2015). Depressive symptoms in third-grade teachers: Relations to classroom quality and student achievement. *Child Development, 86*, 945–954. doi:10.1111/cdev.12344

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. doi:10.1006/cogp.1999.0734

Moon, B. (2007). School-based teacher development in Sub-Saharan Africa: Building a new research agenda. *Curriculum Journal, 18*, 355–371. doi:10.1080/09585170701590007

Neuman, M. J., Josephson, K., & Chua, P. G. (2015). *A review of the literature: Early childhood care and education (ECCE) personnel in low- and middle-income countries* (Early Childhood Care and Education

Working Papers Series). Paris: UNESCO.

Obradović, J., Finch, J. E., Portilla, X. A., Rasheed, M. A., Tirado-Strayer, N., & Yousafzai, A. K. (2019). Early executive functioning in a global context: Developmental continuity and family protective factors. *Developmental Science*. doi:10.1111/desc.12795

Ozler, B., Fernald, L. C. H., Kariger, P., McConnell, C., Neuman, M., & Fraga, E. (2018). Combining preschool teacher training with parenting education: A cluster-randomized con- trolled trial. *Journal of Development Economics,133*, 448–467. doi:10.1016/j.jdeveco.2018.04.004

Phillips, D., Lipsey, M. W., Dodge, K. A., Haskins, R., Bassok, D., Burchinal, M.R., ...Weiland, C. (2017). *Puzzling it out: The current state of scientific knowledge on pre-Kindergarten effects. A consensus statement*. Washington, DC: Brookings Institution.

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. doi:10.1207/s1532480xads0903_2.

Pisani, L., Borisova, I., & Dowd, A. J. (2018). Developing and validating the International Development and Early Learning Assessment (IDELA). *International Journal of Educational Research*, *91*, 1–15. doi:10.1016/j.ijer.2018.06.007

Puma, M., Bell, S., Cook, R., Heid, C., Broene, P., Jenkins, F., ... & Downer, J. (2012). *Third grade follow-up to the Head Start Impact Study: Final report* (OPRE Report 2012-45). Washington, DC: Administration for Children and Families. Retrieved fromhttps://files.eric.ed.gov/fulltext/ED539264.pdf

Raver, C. C. (2003). Young children's emotional development and school readiness. *Social Policy Report*, *16*(3), 3–19.

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. doi:10.1111/j.1467-8624.2010.01561.x

RTI International. (2009a). *Early Grade Reading Assessment (EGRA) toolkit*. Prepared for the World Bank,

Office of Human Development. Research Triangle Park, NC: Author.

RTI International. (2009b). *Early Grade Mathematics Assessment (EGMA): A conceptual framework based on mathematics skills development in children*. Research Triangle Park, NC: Author.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: Wiley.

Sandefur, J. (2016). *Internationally comparable mathematics scores for fourteen African countries* (Working Paper No. 444). Washington, DC: Center for Global Development.

Seidman, E., Kim, S., Raza, M., Ishihara, M., & Halpin, P. F. (2018). Assessment of pedagogical practices and processes in low and middle income countries: Findings from secondary school classrooms in Uganda. *Teaching and Teacher Education*, *71,* 283–296. doi:10.1016/j.tate.2017.12.017

Shonkoff, J. P., & Phillips, D. (2000). *From neurons to neighborhoods: The science of early child development*. Washington, D.C.: National Academy Press.

Skaalvik, E. M., & Skaalvik, S. (2010). Teacher self-efficacy and teacher burnout: A study of relations. *Teaching and Teacher Education, 26*(4), 1059–1069. https://doi.org/10.1016/j.tate.2009.11.001

Smith-Donald, R., Raver, C. C., Hayes, T., & Richardson, B. (2007). Preliminary construct and concurrent validity of preschool self-regulation assessment (PSRA) for field-based research. *Early Childhood Research Quarterly, 22*, 173–187. doi:10.1016/j.ecresq.2007.01.002

Snow, C. E., & Van Hemel, S. B. (2008). *Early childhood assessment: Why, what and how*. Washington, D.C: The National Academy Press.

Swain, W. A., Springer, M. G., & Hofer, K. G. (2015). Early grade teacher effectiveness and pre-K effect persistence: Evidence from Tennessee. *AERA Open, 1*(4). doi:10.1177/2332858415612751

UNESCO. (2015). *Education for all global monitoring report regional overview: Sub-Saharan Africa*. Paris: UNESCO.

United Nations. (2015). *Transforming our world: The 2030 agenda for sustainable development*. New York: United Nations.

Vandell, D. L., Burchinal, M., & Pierce, K. M. (2016). Early child care and adolescent functioning at the

end of high school: Results from the NICHD Study of Early Child Care and Youth Development. *Developmental Psychology, 52*, 1634–1645. doi:10.1037/dev0000169

Volunteer Services Organization. (2002). *What makes teachers tick? A policy research report on teachers' motivation in developing countries*. Washington, DC: Author.

What Works Clearinghouse, Institute of Education Sciences. (2017). *What Works Clearinghouse standards handbook version 4.0*. Princeton, NJ: Author.

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. doi: 10.1111/cdev.12099

Willoughby, M. T., Piper, B., Kwayumba, D., & Mccune, M. (2018). Measuring executive function skills in young children in Kenya. *Child Neuropsychology*. doi:10.1080/09297049.2018.1486395

Wolf, S., Aber, J.L., Behrman, J.R, & Peele, M. (2019). Longitudinal causal impacts of preschool teacher training on Ghanaian children's school-readiness: Evidence for persistence and fadeout. *Developmental Science, 22*(5), e12878. https://doi.org/10.1111/desc.12878

Wolf, S., Aber, J.L., Behrman, J.R., & Tsinigo, E. (2019). Experimental impacts of the Quality Preschool for Ghana interventions on teacher professional well-being, classroom quality, and children's school readiness. *Journal of Research on Educational Effectiveness, 12*(1)*,* 10-37. https://doi.org/10.1080/19345747.2018.1517199

Wolf, S., Aber, J. L., Torrente, C., McCoy, M. & Rasheed, D. (2015). Cumulative risk and teacher well-being in the Democratic Republic of the Congo. *Comparative Education Review, 59*(4)*,* 717-742. https://doi.org/10.1086/682902

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. https://doi.org/10.1016/j.ecresq.2017.05.001

Wolf, S., & McCoy, D. C. (2019). Household socioeconomic status and parental investments: Direct and indirect relations with school readiness in Ghana. *Child Development, 90*(1)*,* 160-278. https://doi.org/10.1111/cdev.12899

Wolf, S., Raza, M., Kim, S., Aber, J.L, Behrman, J., & Seidman, E. (2018). Measuring process quality in pre-primary classrooms in Ghana using the Teacher Instructional Practices and Processes System (TIPPS). *Early Childhood Research Quarterly, 45*, 18-30. https://doi.org/10.1016/j.ecresq.2018.05.003

World Bank. (2016). *Total population between the ages 0 to 14.* Washington, DC: World Bank.

Yoshikawa, H., & Kabay, S. (2014). *The evidence base on early childhood care and education in global contexts*. Background paper for the Global Monitoring Report. Paris: UNESCO.

Yoshikawa, H., Leyva, D., Snow, C. E., Trevino, E., Barata, M. C., 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*, 309–322. doi:10.1037/a0038785

Yoshikawa, H., Wuermli, A. J., Raikes, A., Kim, S., & Kabay, S. B. (2018). Toward high-quality early childhood development programs and policies at national scale: Directions for research in global contexts. *Social Policy Report, 31*, 1–36. doi:10.1002/j.2379-3988.2018.tb00091.x

Young, R., & Johnson, D. R. (2015). Handling missing values in longitudinal panel data with multiple imputation. *Journal of Marriage and Family, 77*, 277–294. doi:10.1111/jomf.12144

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols, 1*, 297–301. doi:10.1038/nprot.2006.46

Zelazo, P. D., & Carlson, S. M. (2012). Hot and cool executive function in childhood and adolescence: Development and plasticity. *Child Development Perspectives*, *6*(4), 354–360. doi:10.1111/j.1750-8606.2012.00246.x

Zhai, R., Raver, C. C., & Jones, S. M. (2012). Academic performance of subsequent schools and impacts of early interventions: Evidence from a randomized controlled trial in Head Start settings. *Child and Youth Services Review, 34,* 946–954. doi:10.1016/j.childyouth.2012.01.026

Table 1. *Mean, SD, and inter-item correlations of all study variables*

		M	SD	Range	1	2	3	4	5	6
1	Literacy	0.524	0.177	0.00 - 0.93	1.00					
2	Numeracy	0.470	0.172	0.00 - 0.92	0.71	1.00				
3	Prosocial	0.660	0.148	0.00 – 1.00	0.39	0.39	1.00			
4	Social competence	0.692	0.138	0.00 – 1.00	0.27	0.25	0.11	1.00		
5	Cognitive flexibility	0.548	0.173	0.00 – 1.00	0.25	0.27	0.24	0.10	1.00	
6	Inhibitory control	0.930	0.139	0.27 – 1.00	0.29	0.34	0.22	0.10	0.19	1.00
7	Behavioral regulation	0.923	0.090	0.20 – 1.00	0.42	0.44	0.36	0.16	0.17	0.20

Notes. All variables were rescaled from their original scale to 0-1 and represent the percentage of correct answers in each domain. All correlations statistically significant at $p < .001$.

Table 2. *Proportion of variance in primary school outcomes for children's pre-primary schools and classrooms*

	Child	Class	School
Literacy	0.644	0.122	0.235
Numeracy	0.664	0.315	0.021
Prosocial	0.852	0.060	0.088
Social competence	0.704	0.222	0.074
Cognitive flexibility	0.944	0.011	0.045
Inhibitory control	0.903	0.097	0.000
Behavioral regulation	0.878	0.028	0.094

Table 3. *Treatment impacts on child outcomes 2 years after program implementation*

	b	(SE)	p-value		Effect size (d_{wt})
Literacy					
TT	0.019	(0.010)	0.057	+	0.11
TTPA	-0.007	(0.010)	0.507		-0.04
Numeracy					
TT	0.000	(0.007)	0.968		0.00
TTPA	-0.016	(0.008)	0.036	*	-0.10
Prosocial					
TT	0.012	(0.009)	0.156		0.08
TTPA	0.007	(0.009)	0.450		0.05
Social competence					
TT	0.015	(0.009)	0.097	+	0.11
TTPA	0.005	(0.009)	0.598		0.04
Cognitive flexibility					
TT	0.022	0.009	0.018	*	0.122
TTPA	0.009	0.009	0.339		0.049
Inhibitory control					
TT	0.015	0.009	0.073	+	0.082
TTPA	0.013	0.009	0.153		0.067
Behavioral regulation					
TT	0.015	(0.006)	0.010	**	0.15
TTPA	0.014	(0.006)	0.021	*	0.14

$+ p < .10$; $* p < .05$; $** p < .01$.

Notes. Sample size = 3,435 children nested in 240 schools.
TT = Teacher training condition; TTPA = Teacher training plus parental-awareness condition.
Models include all children sampled at baseline. Children nested in 240 schools from which they were sampled at baseline. All estimates derived from 100 multiply-imputed datasets and include the following control variables from baseline: district, respective baseline score, public vs. private school, KG level, child gender, age, and a set of 5 dummy variables indicating school and child mobility that occurred during the intervention year.

Table 4. *Moderation of impact estimates by classroom quality and teacher burnout*

	Instructional support				Emotional support				Teacher burnout			
	b	(SE)	p-value		b	(SE)	p-value		b	(SE)	p-value	
Literacy												
TT	-0.002	0.037	0.948		-0.032	0.059	0.592		0.053	0.015	0.001	***
TTPA	-0.027	0.039	0.484		-0.052	0.057	0.361		0.016	0.016	0.310	
Moderator	0.006	0.014	0.659		0.010	0.013	0.454		0.010	0.005	0.060	+
TT*Moderator	0.011	0.018	0.535		0.015	0.017	0.388		-0.020	0.007	0.005	**
TTPA*Moderator	0.011	0.019	0.584		0.014	0.017	0.425		-0.013	0.007	0.052	+
Numeracy												
TT	-0.006	0.030	0.831		-0.091	0.050	0.069	+	0.023	0.012	0.053	+
TTPA	-0.018	0.032	0.586		-0.115	0.050	0.021	*	-0.009	0.013	0.481	
Moderator	0.010	0.013	0.411		-0.021	0.012	0.073	+	0.001	0.004	0.756	
TT*Moderator	0.004	0.015	0.811		0.028	0.015	0.063	+	-0.013	0.006	0.018	*
TTPA*Moderator	0.001	0.016	0.951		0.030	0.015	0.042	*	-0.004	0.005	0.480	
Prosocial												
TT	0.046	0.034	0.179		0.018	0.060	0.764		0.020	0.014	0.152	
TTPA	0.009	0.036	0.792		-0.033	0.057	0.561		0.016	0.014	0.245	
Moderator	-0.009	0.015	0.548		-0.007	0.014	0.638		-0.003	0.005	0.505	
TT*Moderator	-0.017	0.017	0.304		-0.002	0.018	0.929		-0.004	0.006	0.479	
TTPA*Moderator	-0.002	0.018	0.926		0.012	0.017	0.479		-0.005	0.006	0.430	
Social competence												
TT	0.003	0.035	0.935		-0.023	0.060	0.697		0.030	0.014	0.033	*
TTPA	-0.031	0.035	0.371		-0.032	0.058	0.586		0.016	0.014	0.257	
Moderator	-0.013	0.014	0.351		0.005	0.013	0.693		-0.001	0.005	0.820	
TT*Moderator	0.007	0.017	0.701		0.012	0.018	0.505		-0.008	0.006	0.210	
TTPA*Moderator	0.019	0.017	0.269		0.011	0.017	0.508		-0.005	0.006	0.420	
Cognitive flexibility												
TT	0.019	0.043	0.660		0.021	0.066	0.751		0.030	0.016	0.053	+
TTPA	0.005	0.042	0.914		-0.047	0.063	0.456		0.017	0.016	0.279	
Moderator	-0.014	0.019	0.914		-0.006	0.016	0.693		0.001	0.006	0.910	
TT*Moderator	0.000	0.021	0.995		0.000	0.020	0.986		-0.006	0.008	0.413	
TTPA*Moderator	0.001	0.021	0.975		0.016	0.019	0.391		-0.006	0.008	0.424	
Inhibitory control												
TT	0.035	0.040	0.384		0.011	0.065	0.865		0.022	0.016	0.143	
TTPA	0.029	0.041	0.482		-0.034	0.064	0.594		0.016	0.016	0.320	
Moderator	-0.012	0.022	0.595		0.008	0.020	0.673		-0.006	0.008	0.415	
TT*Moderator	-0.012	0.020	0.553		0.000	0.019	0.991		-0.006	0.008	0.404	
TTPA*Moderator	-0.010	0.020	0.624		0.013	0.019	0.490		-0.003	0.007	0.731	
Behavioral regulation												
TT	0.004	0.025	0.859		0.044	0.041	0.274		0.027	0.010	0.006	**
TTPA	0.001	0.026	0.961		0.030	0.038	0.427		0.025	0.010	0.014	*
Moderator	-0.012	0.013	0.365		0.016	0.012	0.175		0.003	0.005	0.567	
TT*Moderator	0.005	0.012	0.680		-0.009	0.012	0.451		-0.007	0.005	0.135	
TTPA*Moderator	0.006	0.013	0.635		-0.005	0.011	0.650		-0.006	0.005	0.161	

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

Notes. Sample size = 3,435 children nested in 240 schools.
TT = Teacher training condition; TTPA = Teacher training plus parental-awareness condition.
Models include all children sampled at baseline. Children nested in schools from which they were sampled at baseline. All estimates derived from 100 multiply imputed datasets and include the following control variables from baseline: district, respective baseline score, public vs. private school, KG level, child gender, age, and a set of 5 dummy variables indicating school and child mobility that occurred during the intervention year.
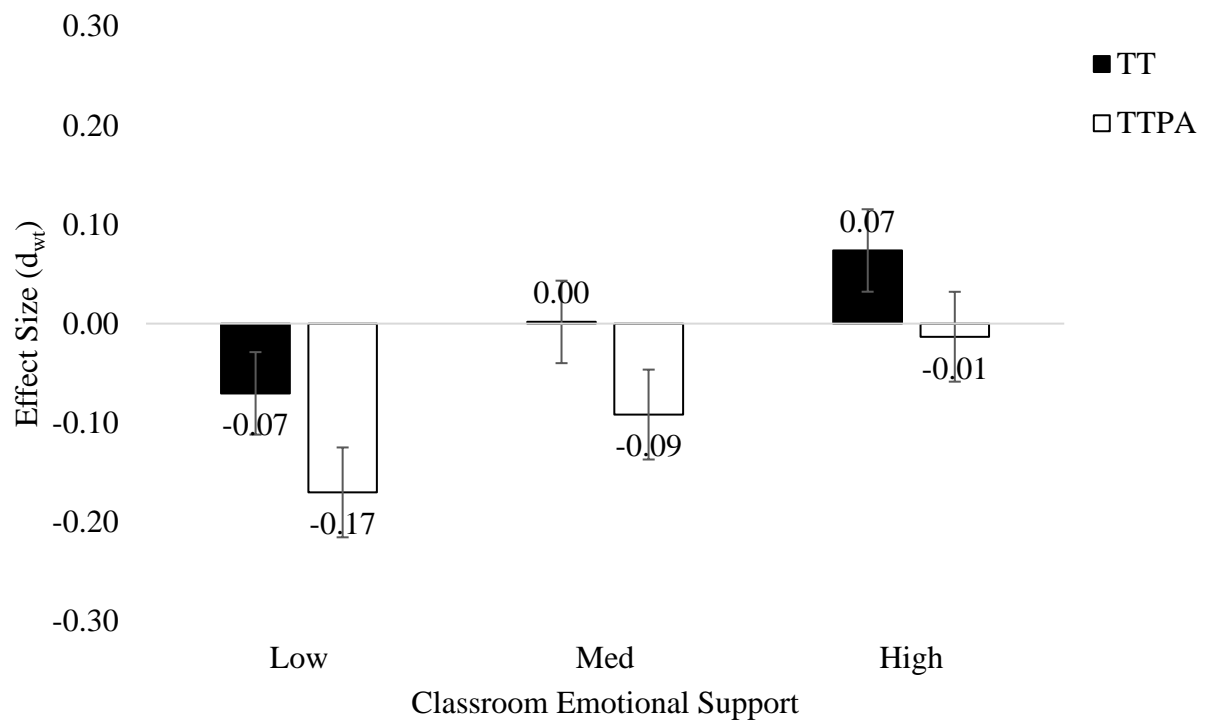
Figure 1. *Impact variation on numeracy by primary classroom emotional support*

Figure 2. *Impact variation in literacy by primary teacher burnout level*
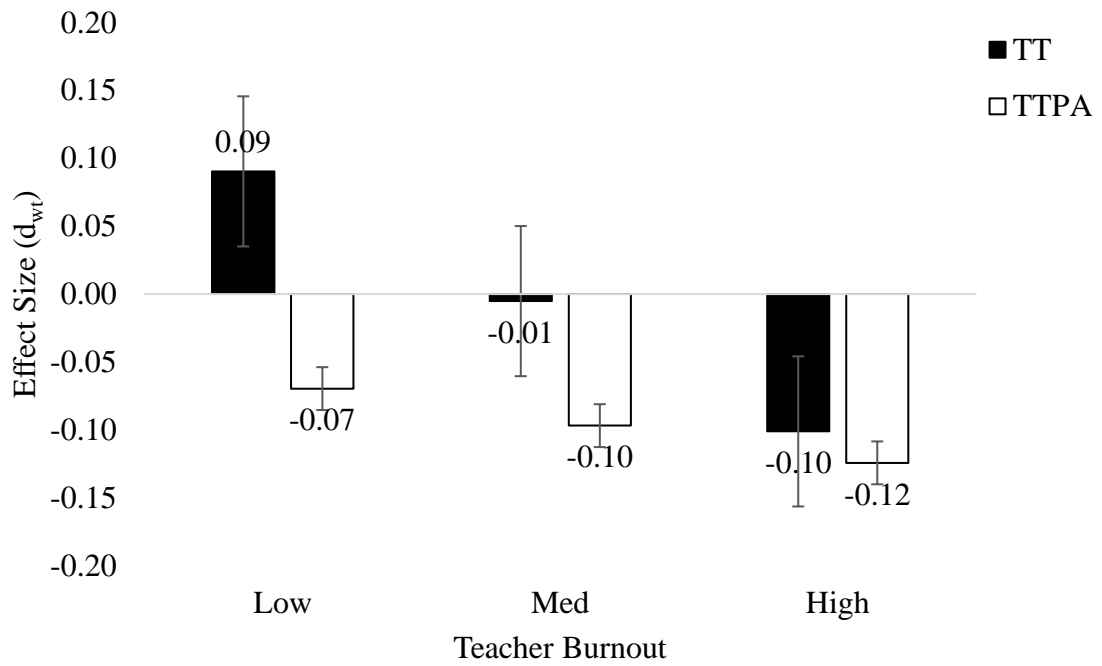

Figure 3. *Impact variation in numeracy by primary teacher burnout level*