



Examining sustained impacts of two teacher professional development programs on professional well-being and classroom practices

Sharon Wolf^{a,*}, Morgan E. Peele^b

^a Graduate School of Education, University of Pennsylvania, 3700 Walnut Street, Philadelphia, PA, 19104, USA

^b Population Studies Center, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA, 19104, USA

HIGHLIGHTS

- Examine year 2 impacts of two teacher professional development programs on teacher professional well-being and classroom practice.
- Most gains observed during the program year faded out.
- Sustained impacts on implementation of training activities were one-third the size of the previous year.
- Parental-awareness meetings had counter-acting effects with negative long-term impacts on one domain of classroom practice.

ARTICLE INFO

Article history:

Received 15 September 2018

Received in revised form

18 June 2019

Accepted 8 July 2019

Available online xxx

Keywords:

Randomized control trial
Teacher professional development
Classroom quality
Ghana
Sub-Saharan Africa

ABSTRACT

We examine second-year impacts of a 1-year pre-primary teacher training and coaching program, delivered with and without parental-awareness meetings, evaluated with a school-level randomized trial. Outcomes included teachers' professional well-being and classroom practices. Most gains observed during the program year faded out. However, there were small sustained impacts on the implementation of training activities. Counteracting effects of the parental-awareness meetings were evident, with suggestive evidence that teachers in this treatment group displayed lower scores on one dimension of classroom quality—Supporting Student Expression—compared to controls. Implications for professional development and educational quality are discussed.

© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

Developing the teacher workforce, including training teachers in ways that improve their pedagogical approaches, is one of the most effective ways to increase student learning (Conn, 2017; McEwan, 2015). Teacher professional development (TPD) programs describe a broad array of approaches to improving teaching practice, and evaluations of such programs often benchmark success based on whether or not programs improve student learning outcomes. In an effort to understand pathways of change (Seidman & Tseng, 2011), some evaluation studies also examine changes at the classroom level that may underlie improvements in student outcomes (e.g., Wolf, Turner, Jukes, & Dubeck, 2018). Both student and

classroom-level outcomes are generally measured only during the period in which a program is implemented.

In sub-Saharan African countries, where resources are often very limited and absolute learning levels are comparatively low (e.g., Sandefur, 2016), understanding the long-term impacts of TPD interventions on teaching practice is of great value. As teachers are the key driver of student learning outcomes (Hanushek, 2002; Rockoff, 2004), it is important to understand whether investments in TPD programs will have lasting gains, or whether more frequent and ongoing training is needed to truly transform teaching practice, particularly in low-resource contexts. While it may be assumed that effective programs will lead to sustained improvements in teaching, it is equally possible that teachers revert to old practices when programs end.

If investments in TPD can transform education systems in a sustained way, it is critical to examine longer term post-intervention impacts. Do TPD efforts need to be ongoing, or are

* Corresponding author.

E-mail address: wolfs@upenn.edu (S. Wolf).

one-time interventions enough to transform teacher practice? While issues of “persistence” and “fadeout” have received attention in the literature on early childhood educational interventions, the focus has been on lasting impacts on student outcomes (e.g., Bailey, Duncan, Odgers & Yu, 2017; Jenkins et al., 2018). Similar considerations for teachers have received less attention, especially at the pre-primary level, although two studies from the United States provide some important insights. Allen and colleagues examined how a web-mediated coaching program for teachers, My Teaching Partner, improved teacher–student interactions and student achievement over a 2-year period in secondary schools (Allen, Pianta, Gregory, Mikami, & Lun, 2011). The authors hypothesized that it would take time for changes in teaching practice to accumulate. At the end of the school year in which the intervention was delivered, there were no impacts on student achievement. The following year, however, with a new class of students and no further coaching of the teacher, there was a significant impact on student achievement, which was mediated by changes in observed teacher–student interactions. A second recent study of the Incredible Years intervention in rural and semirural preschools found that positive changes on classroom climate observed during the implementation year ($d = 0.45$) were not sustained into the next school year. In fact, there were no statistically significant intervention effects for any of the teacher outcomes 1 year later, when teachers were observed with a new group of students (Murray, Rabiner, Kuhn, Pan, & Sabet, 2018).

These studies raise important issues related to the structure of TPD programs and workforce development. We build on these findings by examining sustained changes to teaching practices in pre-primary schools in a very different educational context—namely, peri-urban and semi-rural communities of Ghana. Moreover, we examine the impacts of a teacher training and coaching program delivered with and without supplemental parental-awareness meetings. These meetings consisted of three components: making parents aware of the importance of play-based learning in early childhood education, discussing the role of parents in their children’s learning, and encouraging parent–teacher and parent–school communication. Therefore, in this study we address two sets of critical questions related to longer-term impacts of TPD programs on teachers. First, does a teacher-training and coaching program that had positive short-run effects (i.e., over 1 school year) also promote sustained changes in teaching outcomes the following school year? And do impacts from a supplemental parental-awareness program that had counteracting effects on some aspects of teaching quality in the program-implementation year (Wolf, Aber, Behrman, & Tsinigo, 2019) persist in the following school year? If program effects fade out when a successful program ends, teachers may need ongoing support to sustain gains and truly transform their practice. Second, do sustained impacts differ by teacher characteristics (educational level and years of experience) and school sector (public vs. private)? This has implications for how to target TPD programs, in particular ongoing support, if resources are limited. To address these questions, we use data from an experimental evaluation of a pre-primary teacher training and coaching program, tested with and without parental-awareness meetings, through a randomized control trial in Ghana.

1.1. *Effective teacher professional development programs*

In the past 2 decades there has been a rapid rise in impact evaluations in the field of education. A recent synthesis of 223 rigorous evaluations of in-service training educational initiatives in low- and middle-income countries concluded that low-skilled teachers need specific guidance to reach minimally acceptable

levels of instruction (Ganimian & Murnane, 2016). Another meta-analysis focused on studies in sub-Saharan Africa and found that programs that successfully altered classroom instructional techniques had an effect size of approximately 0.30 standard deviations greater on student learning outcomes than all other types of programs combined (Conn, 2017). These findings point to the importance of not only training teachers in instructional content but in helping them improve their daily interactions with students. In low-resourced countries, in particular, efforts to improve teaching practice are not successful without specifically guided instruction (Ganimian & Murnane, 2016).

These findings parallel a robust and growing evidence base on successful TPD programs in high-income countries, showing the effectiveness of an outside expert providing frequent ongoing support to teachers inside and outside of the classroom (Kraft, Blazar, & Hogan, 2018). For example, interventions designed to provide intensive, individualized coaching to teachers have shown improved teacher–child interactions, especially in high-poverty classrooms (Brown, Jones, LaRusso, & Aber, 2010; Raver et al., 2011) and substantial gains in student achievement (Allen et al., 2011). In their meta-analysis, Kraft et al. (2018) defined coaching programs broadly as all in-service professional development programs where coaches observe teachers in the classroom and provide feedback to help them improve. While coaching fits under the broader umbrella of TPD, Kraft and colleagues see it as distinct from most program offerings, which still consist of short-term and generalized workshops. Interestingly, 51 of the 60 studies reviewed included teachers from kindergarten or elementary school, suggesting that early school grades may be a period when coaching has the potential to be most effective. While in this article we refer to TPD interventions broadly, teacher training coupled with coaching is more likely to yield sustained gains than other programs, based on the literature to date.

Efforts to successfully engage parents in children’s pre-primary education have much less rigorous evidence. The large majority of parenting interventions in early childhood development focus on parenting practices in the first few years of a child’s life. The few studies on the preschool period and children’s school readiness focus on parenting practices and child outcomes (see Welsh, Bierman, & Mathis, 2014 for a review). To our knowledge, no studies have considered how parental engagement interventions affect teacher outcomes and educational quality in the shorter or longer term.

1.2. *Teaching in low-resource settings*

Despite remarkable progress in increasing access to and enrollment in school across the region, there is growing evidence that schoolchildren in sub-Saharan Africa (SSA) are not learning much (e.g., Pritchett, 2013; Sandefur, 2016). This has led governments and global educational goals to change their focus from educational access to educational quality and learning outcomes (e.g., United Nations, 2015). Yet teachers in SSA face many challenges central to quality education that are yet to be addressed systematically, including increasing workloads due to increased student enrollment, low and infrequent teacher remuneration, lack of professional recognition, lack of accountability, and lack of voice (Bennell & Akyeampong, 2007). These challenges have consequences for teaching practices (e.g., Wolf, Aber, Torrente, McCoy, & Rasheed, 2015), as well as for teacher retention (Osei, 2006) and attendance (Bennell & Akyeampong, 2007; Chaudhury, Hammer, Kremer, Muralidharan, & Rogers, 2006), with some scholars pointing to an underlying “motivation crisis” among teachers (Bennell & Akyeampong, 2007; Moon, 2007). Related are other key elements of professional well-being including burnout and work stress (Maslach, Jackson, & Leiter, 1996) and job dissatisfaction

(Zigarelli, 1996), both linked to teacher turnover (e.g., Vahey, Aiken, Sloane, Clarke, & Vargas, 2004). These issues are important to consider in any TPD initiative, particularly when teachers are working under challenging conditions.

Despite numerous programs and interventions to increase teacher effectiveness in the SSA region and in other low- and middle-income countries, many of these have not successfully improved learning outcomes. Those that have generally yielded small impacts on students (see McEwan, 2015, for a meta-analysis). Nearly all of these studies considered impacts during the implementation year only. Were impacts on improved classroom quality sustained the following year? To ensure that such programs will successively lead to transformational change in education systems, it is necessary to examine whether there are lasting impacts of TPD interventions on both professional well-being and changes to pedagogical practices.

1.3. Pre-primary education in Ghana

In 2007, Ghana became one of the first countries in SSA to extend 2 years of pre-primary schooling (i.e., kindergarten for 4-year-olds [KG1] and 5-year-olds [KG2]) as part of its free universal basic educational system. Ghana has among the highest ECE net enrollment rates in Africa at 75% (UNESCO, 2015). Similar to the primary education system (e.g., Ghana Ministry of Education, 2014), reports have concluded that classroom quality and learning outcomes are low among Ghanaian kindergarteners. Given rapid expansion in the demand for kindergarten, a large number of teachers have entered the profession with inadequate training (Ghana Education Service, 2012).

As expansion has coincided with rapid internal migration and urbanization (Central Intelligence Agency, 2016), demand for ECE has expanded, particularly in peri-urban communities (i.e., towns and settlements that surround large cities). The private sector has grown significantly and helped fill a large gap in communities where public preschools have been slower to open (Bidwell & Watine, 2014; Lewis, 2013). The public and private sector teacher workforce differ. For example, there are no requirements for teacher credentials and training in the private sector, while teachers in the public sector are required to have a Diploma in Basic Education obtained from an approved college of education (Asare & Nti, 2014). As a result, private sector teachers are younger and have less formal education (Wolf, Raza et al., 2018). Public sector teachers are also classified as civil servants, and thus receive a guaranteed remuneration level as well as job security, while private sector teachers are generally paid much smaller amounts with no guaranteed job security (Osei, 2006). Nonetheless, children who attend private schools, even low-fee private schools, are more advantaged and have higher school readiness skills (Pesando, Wolf, Behrman, & Tsinigo, 2019; Wamalwa & Burns, 2018).

Examining if and how sustained changes resulting from TPD training differ based on teacher characteristics, such as educational level and years of experience, and private versus public schools, may help to reveal pathways for programs to have successful and sustained impacts. These specific characteristics are central issues for the teacher workforce in Ghana, and in other countries facing a similar growth in private sector education (e.g., Zuilkowski, Piper, Ong'ele, & Kiminza, 2018; Srivastava, 2013).

1.4. The Quality Preschool for Ghana programs

The Quality Preschool for Ghana (QP4G) project aimed to build capacity and support for implementation of the 2004 kindergarten curriculum (Republic of Ghana, 2004) and to enhance the quality of KG education in both public and private schools. The goal was to

develop and rigorously evaluate a scalable model of transformational teacher training to provide high-quality kindergarten instruction to children, and to test the benefits of engaging parents via an awareness campaign designed to align parental expectations with the pedagogy teachers were trained on.

The *teacher-training program* (TT treatment condition) included training workshops and in-classroom coaching administered by trained district-government ECE coordinators led by professional teacher trainers. The training focused on integrating play- and activity-based, child-centered teaching practices into the teaching of instructional content, with opportunities to practice the techniques and create teaching and learning materials to implement activities in the classroom. The coaching visits occurred two times per term (six total visits over the school year) and were implemented by district-government ECE coordinators. These focused on practical ways that teachers could integrate lessons from the training in their teaching, including positive classroom management, assessment and planning, and integrating play-based activities into literacy and math lessons. Teachers were observed for 1 h, followed by debriefing sessions in which teachers reflected on their practice and were provided with feedback on what they did well as well as areas for improvement.

Three *parental-awareness meetings* were administered through school Parent-Teacher Associations (PTAs) over the school year and offered to all parents with kindergarten children. The meetings were implemented at the school and administered by the same district-government ECE coordinators. At each meeting, ECE coordinators screened one of three videos developed for the intervention and led a discussion about the key messages with the parents. 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. The videos and discussions aimed to increase parental involvement at home and in school and increase communication between parents and teachers. It was anticipated that the combined teacher-training and parental-awareness meetings (TPPA treatment condition) would lead to larger impacts than the teacher-training meetings alone, as the parental-awareness meetings would improve parents' understanding of the value of play-based learning, strengthen parent-teacher communication, and increase parental involvement and engagement with teachers.

The programs were designed to improve children's school readiness skills through two primary mechanisms: improved teacher professional well-being and improved classroom quality (see Wolf et al., 2019). These mechanisms are the focus of this study.

1.5. Findings on QP4G impacts on teachers during the program year

An initial study tested the effectiveness of the QP4G teacher-training and coaching program with kindergarten classrooms in private and public schools to improve (a) teacher professional well-being, (b) classroom quality, and (c) children's school readiness skills over 1 school year (Wolf et al., 2019). We summarize key 1-year average treatment effects on teacher outcomes as background for the present study, which are presented in Appendix Table 1.

Regarding teacher professional well-being, there were no impacts on motivation or job satisfaction, but there were reductions in teacher burnout in both the TT ($d_{wt} = -0.40$, $p < .01$) and TPPA ($d_{wt} = -0.59$, $p < .05$) conditions.¹ Additionally, the TT condition

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

impacted teacher turnover, reducing odds that teachers left KG classrooms by the third term of the school year by 58% ($OR = 0.42$, $p < .05$). Similar patterns in the TTPA condition were marginally statistically significant ($OR = -0.53$, $p = .073$). Regarding classroom quality, QP4G increased the number of developmentally appropriate activities teachers used in classrooms in both treatment conditions by similar magnitudes ($d_{wt} = 0.56$, $p < .001$ in TT and 0.61 , $p < .001$ in TTPA). There were no impacts of either treatment condition on teachers' facilitation of deeper learning, but both treatment conditions increased levels of emotional support and positive behavior management observed in the classroom ($d_{wt} = 0.65$, $p < .01$ in TT and 0.66 , $p < .01$ in TTPA). In addition, while the TT condition increased levels of supporting student expression in classrooms ($d_{wt} = 0.52$, $p < .05$), there were no impacts of this dimension of quality in the TTPA treatment arm. These results suggest that the addition of the parental-awareness meetings inhibited some changes in teaching practices.

1.6. The current study

We collected 1-year follow-up longitudinal data on the teachers in the QP4G study, using the same measures as described above for teacher professional well-being and observed classroom quality. Such longitudinal data on both *teachers and classrooms* is rare, particularly in SSA, providing important insights into the potential of TPD efforts to create sustained change in educational systems. The current study goes beyond the initial QP4G study summarized above by addressing two primary questions: (1) Did the short-term intervention-year impacts on teacher well-being and classroom quality persist 1 year after teachers were exposed to QP4G? And, (2) were there differential sustained impacts by baseline teacher (i.e., educational level and years of experience) and school (i.e., public vs. private) characteristics?

2. Methods

The implementation and first-year evaluation of the QP4G intervention occurred between September 2015 and June 2016. The research design was a cluster randomized trial, where 240 schools were randomly assigned to one of the three treatment arms noted above: (a) TT condition: 82 schools, (b) TTPA condition: 79 schools, and (c) control group: 79 schools. The trial was preregistered in the American Economic Associations' registry for randomized controlled trials (RCT ID: AEARCTR-0000704).

Randomization was stratified by district and sector (private and public) to TT, TTPA, or control conditions. Six of the 16 districts in the Greater Accra region were selected. These districts were rated as the most disadvantaged districts in 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) that were within a 2-h drive from Accra (for teachers to be able to attend the training in Accra).

The Greater Accra Region is the most developed part of Ghana, has the smallest proportion of socioeconomically disadvantaged citizens of all the regions, and is rife with ethnic diversity (Owusu & Agyei-Mensah, 2011). According to the 2012 Population and Housing Census, in Ghana the Akans are the predominant ethnic group in Ghana (47.5%), followed by the Mole-Dagbani (16.6%), the Ewe (13.9%), the Ga-Dangme (7.4%), and other groups (14.6%). In the Greater Accra Region specifically, approximately 39.7% are Akan, 27.4% are Ga-Dangme, 20.1% are Ewe, 5.2% are Mole-Dagbani, and 7.6% are other ethnic groups (Ghana Statistical Service, 2013). Importantly, there is tremendous variation in socioeconomic status across districts and neighborhoods. For example, of the six districts

included in this study (Ga South, Adenta, Ledzokuku-Krowor, Ga Central, La Nkwantanang-Madina, and Ga West), the ranking on "disadvantage" ranged from 93 to 187 (average of 139, with higher numbers indicating more disadvantage) out of 216 districts in the country (UNICEF, 2015).

2.1. Power analysis

A sample size of 160 schools (for two-way comparisons) where there were two teachers per classroom per school was assumed. With 80% power at the 5% significance level, and assuming an ICC of 0.10 for teacher outcomes, this was sufficient to detect effect sizes of 0.33.

2.2. Sampling and data collection procedures

School sample. All schools in the six districts were identified using the Ghana Education Service Educational Management Information System (GES-EMIS) database, which lists all registered schools in the country. Eligible schools had to be registered with the government and have at least one KG class. Eligible schools were then randomly sampled within each district and by public and private school. 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 relative to the total for all districts. The sampling and randomization process is shown in Fig. 1.

Teacher sample. All KG teachers in schools selected for the evaluation were invited to participate in the training. The majority of schools had two KG teachers, though the range was from one to five. If there were more than two KG teachers in the school, two teachers were randomly sampled per school for the evaluation (one from KG1 and one from KG2). Thirty-six schools only had one KG teacher, and in this case the one teacher was sampled. The final sample included 444 teachers.

2.3. Measures

The school year in Ghana begins in September and ends in July. The implementation and first-year evaluation of the QP4G intervention occurred between September 2015 (baseline) and June 2016 (follow-up one). A second follow-up occurred in May–June 2017. All data presented in this study were collected in September (baseline) and May–June 2017 (second follow-up).

Teacher professional well-being. Teachers answered a survey in English (the language of Ghana's education system). Items were selected from existing scales and were pilot tested. Factors were derived through exploratory factor analyses conducted with the baseline data.

Motivation. Teachers' motivation was measured using five items adapted from Bennell and Akyeampong (2007) as reported in Wolf et al. (2015). Items were answered on the following scale: 1 = false, 2 = mostly false, 3 = sometimes, 4 = mostly true, 5 = true. Items included "I am motivated to help children develop well socially (i.e., behave well, get along with peers, cooperate)" and "I am motivated to help children learn math" ($M = 4.6$, $SD = 0.59$, $\alpha = 0.69$).

Burnout. Teacher burnout was measured using 11 items from the Maslach Burnout Inventory (Maslach et al., 1996). Items asked teachers to use a scale from 1 ("never") to 7 ("every day") to indicate, for instance, how often they have felt "emotionally drained from my work," "fatigued when I get up in the morning and have to face another day on the job," and "burned out from my work" ($M = 2.09$, $SD = 0.92$, $\alpha = .78$).

Job satisfaction. Teacher's job satisfaction was measured using

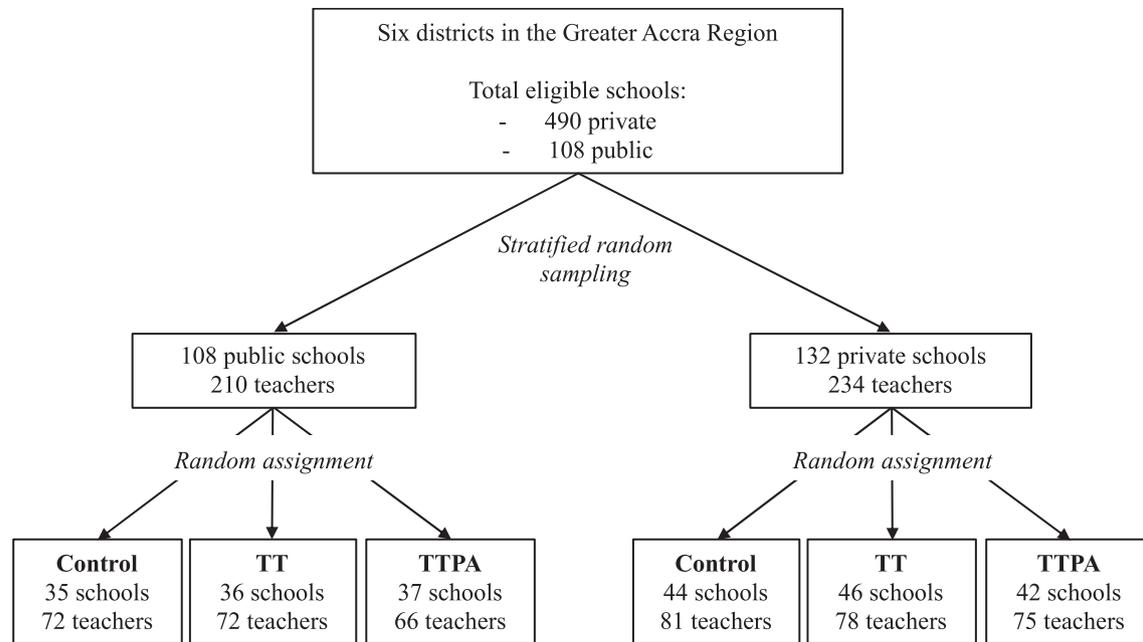


Fig. 1. Sampling at baseline and randomization of schools. Notes. TT = Teacher training condition; TTPA = Teacher training plus parent awareness meetings condition. Total sample includes 240 schools and 444 teachers.

six items adapted from [Bennell and Akyeampong \(2007\)](#) as reported in [Wolf et al. \(2015\)](#). Items were answered on the following scale: 1 = true, 2 = somewhat true, 3 = somewhat false, 4 = false. Sample items include “I am satisfied with my job at this school,” “I want to transfer to another school,” and “Other teachers are satisfied with their decision to be a teacher in this school.” Responses to each item were coded so that higher scores indicated higher job satisfaction ($M = 1.96$, $SD = 0.68$, $\alpha = 0.75$).

Classroom outcomes. All teachers were videotaped teaching a lesson in their classrooms for 30–60 min at each wave. Videos were coded with two instruments: an implementation fidelity checklist and a tool to assess the quality of teacher–child interactions.

Fidelity of implementation/number of training activities. We created a checklist of 15 activities that were explicitly covered in the teacher training related to behavior management and instructional practice to assess the extent to which teachers were using activities from the training in their classrooms. Each practice was coded as either present in the video (a score of 1) or absent in the video (a score of 0). Items included: “Teacher praises children for positive behavior,” “Teacher threatens children with or used a cane on children at least once (reverse coded),” “Teacher explicitly reminds children of the class rules,” “Teacher uses a signal to gain children’s attention (e.g., drum beat, song, bell),” “Children are seated in a way that children can see each other’s faces (e.g., in a circle, or tables together in groups),” “Teacher uses one or multiple songs to facilitate learning at some point in the lesson,” and “There is an activity that facilitated the lesson objectives that involved manipulation of materials” ($M = 3.46$, $SD = 1.63$).

Teacher–child interaction quality. All videos were coded using the Teacher Instructional Practices and Processes System (TIPPS; [Seidman, Kim, Raza, Ishihara, & Halpin, 2018](#)). 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 low- and middle-income countries. We used the TIPPS–Early Childhood Education version and made minor adaptations for use in Ghana (e.g., referring to pupils as children, as is common in Ghanaian kindergarten settings). More information about the

assessment tool can be obtained by referring to [Seidman et al. \(2018\)](#).

Based on an analysis described in detail in [Wolf, Raza et al. \(2018\)](#), we grouped TIPPS items into three factors: *Facilitating Deeper Learning* (three items: connects lesson to teaching objectives; provides specific, high-quality feedback; and uses scaffolding; $\alpha = 0.50$), *Emotional Support and Behavior Management* (seven items: positive climate; negative climate (reverse scored); sensitivity and responsiveness; tone of voice; positive behavior management; provides consistent routines; and student engagement in class activities; $\alpha = 0.81$), and *Supporting Student Expression* (four items: considers student ideas and interests; encourages students to reason and problem solve; connects lesson to students’ daily lives; and models complex language; $\alpha = .61$). See [Wolf, Raza et al. \(2018\)](#) for details on the analysis and concurrent validity of the three factors in this sample.

Inter-rater reliability. Details on the training process for enumerators for the assessment tool can be found in [Wolf, Raza et al. \(2018\)](#). Inter-rater reliability is essential to establishing that the rater’s data are reflective of a mutually agreed upon standard. To assess inter-rater reliability, a randomly selected 15% of videos collected at baseline were coded by three raters ([Hallgren, 2012](#)). To examine inter-rater reliability (IRR) of a four-category scale with a first-stage dichotomous decision, as is the case in the TIPPS, the commonly used IRR statistics (e.g., Cohen’s kappa; [Cohen, 1960](#)) are not ideal. We follow recommendations by [Seidman et al. \(2018\)](#) to calculate the First-order agreement coefficients (AC1; [Gwet, 2002](#)), which are less biased or vulnerable to change in prevalence issues, assume lower chance agreement, and therefore are more stable ([Heyman, Lorber, Eddy, & West, 2014](#)). Thus, we calculate AC1s for the four-category rating on each of the 19 items. The statistic used for these outcomes—that is, the percent of variance explained by the video itself versus the coders—can be interpreted similar to a Kappa value, where above 0.7 (or 70%) indicates sufficient IRR. We calculated the ICC of the final scores to assess how the partition of variance in scores breaks down into differences in individual raters and shared variance across raters. At the second follow-up, on

average across all 19 items 86.4% of the variance was shared across raters, indicating acceptable IRR.

2.4. Analytic strategy

Baseline equivalency across school, teacher, and child characteristics was established and is described in detail in Wolf and colleagues (2019). The results confirm that randomization successfully yielded three groups equivalent on observable characteristics.

Of the 444 teachers who were recruited for the study at baseline, 348 remained in the first follow-up at the end of the 2015–2016 academic year. Teachers were missing from the follow-up sample for various reasons, including leaving the teaching profession, moving into a school administration position, or switching to teach a grade level above third grade ($N = 63$). At the second follow-up at the end of the 2016–2017 academic year, an additional 26 teachers were not surveyed for the same reasons. Furthermore, 13 teachers refused to participate or were not available, resulting in 70% of the baseline sample ($N = 309$) available at the second follow-up. The 1-year impact analysis found that treatment status was a significant predictor in teacher attrition (30.7% in control, 16.9% in TT, 18.7% in TTPA; $\chi^2(2) = 9.56, p = .008$). Treatment status did not significantly predict teacher attrition in the second year ($\chi^2(2) = 0.15, p = .928$).

Missing data imputation. We used multiple imputation (with Stata's "ice" command) to address missing data on all missing variables, using three rounds of data collection (baseline and follow-up, as well as a second round of follow-up data). While the data are not missing completely at random (MCAR), if variables that strongly predict attrition are incorporated into the missing data strategy, the plausibility of a missing at random (MAR) assumption increases (Young & Johnson, 2015). In other words, the assumptions of MAR have been shown to be robust when a large set of covariates are included in the chained-model estimation. Our imputation approach meets the standards of the What Works Clearinghouse Version 4.0 Standards Handbook (IES, 2017). Using a rich set of teacher demographic and background variables, outcome scores for professional well-being and classroom quality across all waves, and treatment status indicators, we imputed 20 teacher-level datasets. All impact estimates on teacher- and classroom-level data were computed on these 20 datasets (using Stata's "mi estimate" command). Ten teachers refused to have their classrooms videotaped at the first follow-up, as did 12 teachers at the second follow-up, and these data were imputed in addition to the data for teachers who left the sample.

Impact analysis. We employed two-level multilevel modeling in Stata (Version 14.0). While the number of teachers per school was small (i.e., ~2 teachers/school), schools were the unit of randomization, and teachers were sampled within schools. First, we estimated unconditional models to estimate the ICCs, or the proportion of variance in each of the teacher/classroom outcomes attributable to teachers/classrooms and schools at the second follow-up. Second, impact analyses were conducted with a select set of covariates. The multiply-imputed datasets were used in all analyses with Stata's "mi estimate" command, which uses Rubin's combining rules to compute pooled coefficients and standard errors across datasets (Rubin, 1987).

Separate models were fitted to estimate main intervention impacts on (a) teacher professional well-being (i.e., motivation, burnout, and job satisfaction), and (b) classroom quality factors (i.e., fidelity of implementation checklist, facilitating deeper learning, emotional support and behavior management, and supporting student expression). The equation is as follows:

Level 1 (Teacher-level) Model:

$$Y_{ij} = \beta_{0j} + \beta_1 X_{ij} + \beta_2' Z_{ij} + e_{ij}$$

where X_{ij} is the moderating teacher characteristic, and Z_{ij} is a vector of additional teacher covariates (gender, age, education level, years of experience, and baseline score).

Level 2 (School-level) Model:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} TT_j + \gamma_{02} TTPA_j + \gamma_{03}' Z_j + u_{0j}$$

where γ_{00} is the school-level random intercept; Z_j is the vector of school-level covariates (district dummies, private or public status, and four dummy variables for different school mobility scenarios [i.e., three treatment schools combined their separate KG classrooms into one KG1 and one KG2 classroom; one school split into two schools between baseline and follow-up; nine schools started with a combined KG1 and KG2 class, and split into two separate classrooms midyear; and 12 teachers switched to teach a different KG class within the school midyear]); and TT_j is an indicator for schools assigned to the teacher training condition, and $TTPA_j$ an indicator for schools assigned to condition involving teacher training plus parental-awareness meetings.

Third, as a secondary exploratory analysis, we examined whether longer-term intervention impacts varied by teacher characteristics (educational level and years of experience) and by school sector (private vs. public). Impact variation by teacher characteristics was tested by adding a cross-level interaction term between each treatment condition (at level 2) and teacher characteristic (at level 1). Moderation by sector was calculated with an interaction term (at level 2) between school sector ($I = \text{private}, 0 = \text{public}$) and treatment status. The moderating effect is modeled by two cross-level interaction terms between γ_{01} and β_1 and γ_{02} and β_2 .

3. Results

Inter-item correlations for all outcome variables are presented in Table 1, and intra-class correlations for all outcomes are shown in Table 2. Burnout was negatively correlated with both motivation ($r = -0.25$) and job satisfaction ($r = -0.26$), while motivation and job satisfaction were positively correlated ($r = 0.17$). The number of training activities used had small positive correlations with all three dimensions of teacher–child interaction quality ($r = 0.15–0.23$), and the three dimensions of teacher–child interaction quality were positively correlated ($r = 0.15–0.27$). There was no statistically significant correlation between professional well-being and classroom processes.

3.1. Persistent impacts on professional well-being

The first panel in Table 3 displays the results for persistent treatment impacts on teachers' motivation, burnout, and job satisfaction. While the 1-year impacts showed a moderate-large reduction in burnout in both treatment arms, there were no impacts on any of the three outcomes at the second follow-up.

3.2. Persistent impacts on classroom quality

Longer-term impacts on four dimensions of classroom outcomes are displayed in the second panel of Table 3. There were persistent positive impacts on the fidelity of implementation checklist, representing the number of activities from the training that teachers used in their classrooms, with marginally statistically significant impacts in the TT condition ($d_{wte} = 0.21, p = .065$) and reaching statistical significance in the TTPA condition ($d_{wte} = 0.25, p = .037$).

Table 1
Inter-item correlations among all study variables.

		1	2	3	4	5	6
1	Motivation						
2	Burnout	-0.25					
3	Job satisfaction	0.17	-0.26				
4	Number of training activities used	0.08	0.03	-0.01			
5	Facilitating deeper learning	-0.00	0.11	-0.11	0.16		
6	Emotional support & behavior management	-0.02	0.04	0.01	0.15	0.15	
7	Supporting student expression	0.04	0.08	0.01	0.23	0.27	0.24

Note. Bold numbers indicate correlation is statistically significant at $p < .05$.

Table 2
Intra-class correlations for all outcome variables.

	Proportion of variance	
	Teacher	School
Motivation	0.998	0.002
Burnout	0.751	0.249
Job satisfaction	0.698	0.302
Number of training activities used	0.689	0.311
Facilitating deeper learning	0.726	0.274
Emotional support & behavior management	0.730	0.270
Supporting student expression	0.719	0.281

Notably, these effect sizes are about one-third of the effect size found in the first follow-up ($d_{wt} = 0.56$ in TT and 0.61 in TTPA, $p < .001$).

Three dimensions of teacher-child interaction quality were examined. As with the first follow-up, there were no impacts on teachers' facilitation of deeper learning practices at the second follow-up 1 year later. While the first-year follow-up showed gains in teachers' emotional support and behavior management ($d_{wt} = 0.65$ in TT and 0.66 in TTPA, $p < .01$), these were no longer statistically significant at the second follow-up. Finally, there were

Table 3
Year-2 impacts on teacher and classroom outcomes.

	<i>b</i>	(SE)	<i>p</i> -value	d_{wt}
Teacher professional well-being				
Motivation				
TT	-0.005	(0.077)	0.949	-0.019
TTPA	0.018	(0.079)	0.822	0.068
Burnout				
TT	-0.149	(0.153)	0.332	-0.163
TTPA	-0.248	(0.163)	0.130	-0.273
Job satisfaction				
TT	0.019	(0.105)	0.858	0.047
TTPA	-0.029	(0.111)	0.796	-0.072
Classroom outcomes				
Number of training activities used				
TT	0.471	(0.254)	0.065	+ 0.212
TTPA	0.545	(0.261)	0.037	* 0.245
Facilitating deeper learning				
TT	-0.132	(0.103)	0.202	-0.304
TTPA	-0.181	(0.112)	0.108	-0.417
Emotional support & behavior management				
TT	0.043	(0.555)	0.443	0.359
TTPA	0.016	(0.064)	0.805	0.135
Supporting student expression				
TT	-0.144	(0.119)	0.230	-0.342
TTPA	-0.288	(0.118)	0.016	* -0.682

Notes. TT = Teacher training condition; TTPA = Teacher training plus parental-awareness meetings condition. Estimates are computed using observed scores, in two-level models: teachers nested in schools. Effect sizes calculated accounting for the nested model structure (Hedges, 2009). Sample size: 444 teachers nested in 240 schools. All impact estimates computed from 20 multiply imputed datasets. + $p < .10$, * $p < .05$.

no statistically significant impacts on supporting student expression in the TT condition at the second follow-up, but there were statistically significant and moderate-sized negative impacts in the TTPA condition ($d_{wt} = -0.68$, $p = .016$).

3.3. Moderation in persistent impacts by teacher and school characteristics

Finally, we examined whether sustained impacts on professional well-being and teaching practice were moderated by teacher educational level, years of experience, and school sector (private vs. public). There were no significant differences across any of the variables examined with one exception (see Table 4). Specifically, teachers' years of experience moderated sustained impacts of the TT treatment on emotional support and behavior management. Given the overall pattern of insignificant moderation across the three moderators, we do not interpret this as a meaningful effect.

4. Discussion

We examined whether a teacher-development program that improved some elements of teachers' professional well-being and classroom practices during the implementation year led to sustained changes 1 year later. Using a school-level randomized experiment, we found that previously observed reductions in teacher burnout faded out, and there were no differences between treatment and control group teachers across any indicator of professional well-being. Teachers maintained some changes to their teaching practice, specifically in their use of more developmentally appropriate instructional and behavior management activities that were a central focus of the training. Our finding of sustained gains in concrete activities that can easily be integrated into instruction, as opposed to more abstract teaching methods, is in line with a recent literature review on teacher professional development programs that found that a focus on concrete activities made the biggest difference to teachers (Walter & Briggs, 2012). Importantly, the size of impacts was smaller ($d_{wt} = 0.21-0.25$, $p < .07$) compared to the previous year ($d_{wt} \sim 0.60$, $p < .05$), indicating that the level of gains was not maintained.

Changes in teacher-child interaction quality, on the other hand, either faded out completely or were reversed. Previously observed improvements in teachers' use of positive emotional support and behavior management, which were observed in both treatment conditions the previous year, were no longer statistically significant. For the teacher training and coaching treatment arm (TT), gains observed in teachers' support for student expression in the implementation year ($d_{wt} = 0.52$, $p < .05$) faded out the following year. The second treatment arm, which added parental-awareness meetings alongside the teacher training (TTPA), showed that these meetings inhibited the improvements in support for student expression during the implementation year ($d_{wt} = 0.09$, $p = n.s.$). One year later, teachers in this treatment arm had lower levels of

Table 4
Moderation of Year-2 impacts by teacher and school characteristics.

	Education level			Years of experience			Private (vs public)		
	<i>b</i>	(SE)	<i>p</i> -value	<i>b</i>	(SE)	<i>p</i> -value	<i>b</i>	(SE)	<i>p</i> -value
Teacher professional well-being									
Motivation									
TT x Moderator	−0.024	(0.028)	0.399	0.000	(0.001)	0.973	0.010	(0.129)	0.441
TTPA x Moderator	−0.018	(0.029)	0.541	−0.004	(0.010)	0.699	0.067	(0.128)	0.603
Burnout									
TT x Moderator	0.008	(0.055)	0.881	0.001	(0.018)	0.953	−0.363	(0.230)	0.115
TTPA x Moderator	0.006	(0.056)	0.916	−0.003	(0.019)	0.863	−0.266	(0.236)	0.260
Job satisfaction									
TT x Moderator	0.023	(0.036)	0.527	−0.019	(0.013)	0.141	−0.047	(0.162)	0.773
TTPA x Moderator	0.028	(0.038)	0.448	−0.008	(0.012)	0.497	−0.045	(0.170)	0.791
Classroom outcomes									
Number of training activities used									
TT x Moderator	−0.074	(0.095)	0.438	−0.010	(0.032)	0.751	0.101	(0.422)	0.811
TTPA x Moderator	−0.015	(0.093)	0.876	0.010	(0.033)	0.753	0.397	(0.422)	0.347
Facilitating deeper learning									
TT x Moderator	−0.032	(0.038)	0.401	0.005	(0.013)	0.699	0.135	(0.180)	0.453
TTPA x Moderator	−0.007	(0.042)	0.865	−0.005	(0.013)	0.696	−0.056	(0.186)	0.762
Emotional support & behavior management									
TT x Moderator	0.025	(0.021)	0.226	−0.014	(0.007)	0.043	−0.068	(0.097)	0.483
TTPA x Moderator	−0.011	(0.021)	0.592	−0.009	(0.007)	0.174	0.058	(0.087)	0.503
Supporting student expression									
TT x Moderator	−0.044	(0.043)	0.305	−0.005	(0.012)	0.662	0.055	(0.179)	0.760
TTPA x Moderator	−0.030	(0.044)	0.496	−0.010	(0.013)	0.443	−0.036	(0.188)	0.850

Notes. TT = Teacher training condition; TTPA = Teacher training plus parental-awareness meetings condition. Estimates are computed using observed scores, in two-level models: teachers nested in schools. Sample size: 444 teachers nested in 240 schools. All impact estimates computed from 20 multiply imputed datasets. * $p < .05$.

support for student expression than teachers in the control group, and these differences were moderately large ($d_{wt} = -0.68$, $p < .05$).

The counter-acting effects of the parental-awareness meetings are consistent with the first follow-up study, which found that these meetings also counteracted improvements in child outcomes (Wolf et al., 2019). Qualitative follow-up data suggest that engaging parents regarding the new teaching methods may have led some parents to pressure teachers to maintain traditional and familiar teaching methods (see Kabay, Wolf, & Yoshikawa, 2017; Wolf et al., 2019). When the teacher training ended, it is possible that this pressure from parents continued. In the absence of ongoing training and coaching, teachers may have overcompensated in response to parents by reducing their use of practices related to supporting students to express themselves. These specific practices, more than the other changes observed as a result of the training, directly countered teaching methods often used in Ghana. Typically, teacher-led instruction and rote memorization are favored, as children are viewed as “receptacles” that should absorb information rather than active participants in the classroom (Agbenyega, 2017). This traditional view may partially explain why these particular practices were difficult to maintain, though these explanations are only speculative. Importantly, we consider the negative impact as suggestive, rather than conclusive.

There may be other explanations that are related to the study design. For example, given that this was the only statistically significant negative impact among seven outcomes that was detected, it is possible that the result was a Type I error. Alternatively, while baseline equivalence was established (Wolf et al., 2019), it is 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. These interpretations are also limited by the fact that in the second follow-up, the measure of supporting student expression had relatively low internal consistency.

The size of impacts on classroom quality in the first year of the program ($d = 0.52$ – 0.66 ; see Appendix Table 1) are in line with

effect sizes found in other TPD evaluations. Specifically, a recent meta-analysis of quasi-experimental evaluations of ECE TPD programs showed an average effect size of $d = 0.68$ on classroom outcomes during the ECE teacher training year (Egert, Fukkink, & Eckhardt, 2018). Two other meta-analyses focused on child care centers found effect sizes on classroom quality of $d = 0.45$ (Fukkink & Lont, 2007) and $d = 0.35$ – 0.39 (Werner, Linting, Vermeer, & van Ijzendoorn, 2016). For the handful of studies that have considered impacts 1 year after treatment, results show large variation in effect sizes, ranging from -0.05 to 0.60 (Allen et al., 2011; Murray et al., 2018).

One strength of this study is that it includes data on both teacher professional well-being and observed classroom practices 1 year after teachers completed the training program. A synthesis of over 200 rigorous evaluations of in-service training educational initiatives in low- and middle-income countries concluded that resources improve student achievement only if they change children’s interactions at school (Ganimian & Murnane, 2016). Therefore, observing and measuring teacher–child interactions is key to uncovering pathways of change through which TPD programs ultimately improve student outcomes (e.g., Wolf, Turner et al., 2018). In a review of meta-analyses analyzing educational interventions implemented in sub-Saharan Africa, Evans and Popova (2016) concluded that one important ingredient to successful pedagogical interventions was repeated and ongoing trainings. The findings from this study indicate that once ongoing trainings end, teachers struggle to maintain changes to their practice, even if a training was initially successful. If teachers face professional challenges from both inside and outside of school, they might find it even more difficult to maintain these changes (e.g., Wolf et al., 2015).

4.1. Future research

The results of this study inform the provision of high-quality professional development and learning opportunities for teachers, which is among the most important and longstanding

challenges faced by education systems. While the evidence is clear that providing ongoing support to teachers and in-classroom coaching can improve teaching practices in both high-income (Kraft et al., 2018) and low- and middle-income countries (e.g., Wolf, 2018; Wolf et al., 2019; see Popova, Evans, Breeding, & Arancibia, 2018 for a review), high-quality programs have proven difficult to scale and sustain. The results of the Quality Preschool for Ghana program provide no exception.

There are several important areas for future research and practice. First, educational evaluations that hypothesize changes to teachers and classrooms as the key mechanisms to improved student learning should in fact measure these processes. The results from this study shed light on why engaging parents—hypothesized to strengthen impacts on teaching practice and child outcomes—may have counteracted the positive changes to teaching practice and child outcomes as reported in Wolf et al. (2019) by showing how the program led to negative impacts on one dimension of teaching practice. Measuring key dimensions in program theories of change in future studies will lead to a better understanding of the pathways underlying effective (or ineffective) interventions. Second, future studies should consider if changes observed during the TPD training period persist, both for teachers and students, as well as how additional supports can be put in place to ensure teachers sustain gains to their professional well-being and teaching practices. This should particularly be done for programs involving teachers in low-resource settings, where daily structural challenges can hinder teachers' effectiveness and ability to implement new practices. Both research and practice should consider these to be central issues if TPD is to effectively lead to transformational system-level change. Once a larger body of evidence exists, it will be possible to begin to analyze why some programs yield sustained impacts, and others do not.

4.2. Limitations and conclusions

The results from this study must be interpreted in light of its limitations. First, there was significant attrition of the teachers in the sample (about one-third of the baseline sample), particularly between the baseline and first follow-up. The use of multiple imputation and multiple controls probably limits any bias due to attrition. Notably, the impacts on only teachers who remained in the sample yield a similar pattern of results (see Appendix Table 2). Second, the study's sample is limited to six peri-urban and semi-rural districts in the Greater Accra Region in Ghana, with implications for generalizability of the findings. Third, there was relatively low internal consistency for two of the classroom quality domains: facilitating deeper learning ($\alpha = 0.50$) and supporting student expression ($\alpha = 0.61$). Thus, we cannot rule out the role of poor measurement in driving the findings. Finally, due to resource constraints, very little data were collected on the implementation of the parental-awareness training and parents' engagement in and perceptions of this training. It is noteworthy that 1 year after the intervention, teachers in this condition displayed lower levels of quality related to supporting student expression in their classrooms. Because of the lack of implementation data, we can only speculate about these unexpected findings.

Despite these limitations, this study has important implications for the development of the teacher workforce, as significant resources are spent on teacher professional development by governments and international agencies. For example, of 171 World Bank projects with educational components between 2000 and 2012, approximately two-thirds included professional

development to support teachers (Popova et al., 2018). Importantly, most of these programs were implemented over the course of a few months or 1 year at best. It is known that effective programs include sustained training and ongoing support such as coaching (Darling-Hammond, Wei, Andree, Richardson, & Orphanos, 2009; Kraft et al., 2018), but even these programs rarely last beyond 1 academic year. In a study of elementary school teachers involved in a language arts PD program, Cohen, Schult, Brown, and Grossman (2016) found that the duration of professional development was a key predictor of teachers' appropriation of the training practices. Teachers who participated in a single year of the PD, compared to those who participated in 2 years, used the targeted practices less frequently and with lower levels of sophistication.

If providing training to teachers does not lead to sustained improvements in teaching practices and student learning outcomes across multiple years, then these resources may be better spent in other ways. One solution may be to ensure that there is alignment of TPD programming across school years, and that repeated training and ongoing coaching continue for at least 2 academic years. Persistence and fadeout of educational interventions on teachers and classrooms, in addition to students, should become a central issue in global efforts for educational quality and learning for all. More research and input from teachers themselves on how daily work conditions affect their teaching practice are needed to create training structures that lead to sustained changes for both teachers and students.

Appendix Table 1

Year-1 impacts on teacher professional well-being and classroom quality

	<i>b</i>	(SE)	<i>p</i> -value	<i>d_{wt}</i> or OR
Teacher professional well-being				
Motivation				
TT	0.103	(0.068)	0.132	0.345
TTPA	0.033	(0.072)	0.648	0.111
Burnout				
TT	-0.330	(0.155)	0.035	* -0.398
TTPA	-0.492	(0.159)	0.002	** -0.593
Job satisfaction				
TT	0.130	(0.096)	0.175	0.328
TTPA	0.043	(0.100)	0.663	0.108
Teacher turnover ^a				
TT	-0.868	(0.414)	0.036	* 0.420 ^b
TTPA	-0.629	(0.351)	0.073	+ 0.533 ^b
Classroom outcomes				
Fidelity checklist (no. of activities)				
TT	1.310	(0.246)	0.000	*** 0.560
TTPA	1.434	(0.266)	0.000	*** 0.613
Facilitating deeper learning				
TT	-0.045	(0.106)	0.672	-0.113
TTPA	-0.063	(0.116)	0.588	-0.158
Emotional support & behavior management				
TT	0.170	(0.065)	0.010	** 0.647
TTPA	0.172	(0.066)	0.010	** 0.655
Supporting student expression				
TT	0.235	(0.110)	0.033	* 0.524
TTPA	0.042	(0.116)	0.719	0.094

Notes. Table reprinted from Wolf et al., 2019.

TT = Teacher training condition; TTPA = Teacher training plus parental-awareness meetings condition. Estimates are computed using observed scores, in two-level models: teachers nested in schools. Effect sizes calculated accounting for the nested model structure (Hedges, 2009). Sample size: 444 teachers nested in 240 schools. All impact estimates computed from 20 multiply imputed datasets.

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

^a Model estimated with multinomial logistic regression.

^b Odds ratio presented.

Appendix Table 2

Year-2 impacts for teachers present at baseline and second follow-up only

	<i>b</i>	(SE)	<i>p</i> -value	
Teacher professional well-being				
Motivation				
TT	−0.043	0.064	0.505	
TTPA	−0.090	0.064	0.159	
Burnout				
TT	0.048	0.141	0.735	
TTPA	0.209	0.141	0.139	
Job satisfaction				
TT	−0.047	0.114	0.681	
TTPA	0.119	0.114	0.296	
Classroom outcomes				
Number of training activities used				
TT	0.868	0.275	0.002	**
TTPA	0.920	0.280	0.001	***
Facilitating deeper learning				
TT	−0.049	0.121	0.661	
TTPA	−0.169	0.123	0.160	
Emotional support & behavior management				
TT	0.010	0.065	0.876	
TTPA	0.101	0.067	0.131	
Supporting student expression				
TT	−0.171	0.107	0.109	
TTPA	−0.260	0.110	0.017	*

Notes. TT = Teacher training condition; TTPA = Teacher training plus parental-awareness meetings condition. Estimates are computed using observed scores, in two-level models: teachers nested in schools. Effect sizes calculated accounting for the nested model structure (Hedges, 2009). Sample size = 309 teachers in 203 schools.

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

Acknowledgements

This paper reflects contributions from many organizations and individuals. First, we would like to thank J. Lawrence Aber and Jere R. Behrman, co-investigators on the project who provided input on all aspects of the larger study as well as feedback on this article; we are grateful for all of their support. Second, we would like to thank the committed staff and thought partners at Innovations from Poverty Action, including Loic Watine, Bridget Gyamfi, Renaud Comba, Kwabena Asamonah, Henry Atimone, and Edward Tsinigo, and our talented data collection supervisors and enumerators. We also thank the Ghana Education Service for their support of this project. And, we would like to thank the dedicated teachers, children, and families who participated in the study. We thank the UBS Optimus Foundation and the World Bank Strategic Impact Evaluation Fund for funding this project.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.tate.2019.07.003>.

References

- Agbenyega, J. S. (2017). Examining early childhood education system in Ghana: How can Bourdieuan theorization support a transformational approach to pedagogy? In M. Fleer, & B. Oers (Eds.), *International Handbook of early childhood education* (pp. 673–690). Netherlands: Springer. https://doi.org/10.1007/978-94-024-0927-7_32.
- Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science*, *333*(6045), 1034–1037.
- Asare, K. B., & Nti, S. K. (2014). Teacher education in Ghana: A contemporary synopsis and matters arising. *SAGE Open*, *4*(2). <https://doi.org/10.1177/2158244014529781>.
- 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. <https://doi.org/10.1080/19345747.2016.1232459>.
- Bennell, P., & Akyeampong, K. (2007). *Teacher motivation in sub-Saharan Africa and*

- South Asia*. London: DfID.
- Bidwell, K., & Watine, L. (2014). *Exploring early education problems in peri-urban settings in Africa: Final report*. Newhaven, CT: Innovations for Poverty Action.
- Brown, J. L., Jones, S. M., LaRusso, M. D., & Aber, J. L. (2010). Improving classroom quality: Teacher influences and experimental impacts of the 4rs program. *Journal of Educational Psychology*, *102*(1), 153.
- Central Intelligence Agency. (2016). *Ghana. The world factbook*. Retrieved Feb 22 2017 from: <http://www.cia.gov/worldfactbook.us/africa/ghana.html>.
- Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K., & Rogers, F. H. (2006). Missing in action: Teacher and health worker absence in developing countries. *The Journal of Economic Perspectives*, *20*(1), 91–116.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46. <https://doi.org/10.1177/001316446002000104>.
- Cohen, J., Schult, L. C., Brown, L., & Grossman, P. (2016). Leveraging observation tools for instructional improvement: Exploring variability in uptake of ambitious instructional practices. *Teachers College Record*, *118*(11), 1–36.
- Conn, K. M. (2017). Identifying effective education interventions in sub-Saharan Africa: A meta-analysis of impact evaluations. *Review of Educational Research*, *87*(5), 863–898.
- Darling-Hammond, L., Wei, R. C., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession*. Washington, DC: National Staff Development Council.
- Egert, F., Fukkink, R. G., & Eckhardt, A. G. (2018). Impact of in-service professional development programs for early childhood teachers on quality ratings and child outcomes: A meta-analysis. *Review of Educational Research*, *88*(3), 401–433.
- Evans, D. K., & Popova, A. (2016). What really works to improve learning in developing countries? An analysis of divergent findings in systematic reviews. *The World Bank Research Observer*. <https://doi.org/10.1093/wbro/lkw004>.
- Fukkink, R. G., & Lont, A. (2007). Does training matter? A meta-analysis and review of caregiver training studies. *Early Childhood Research Quarterly*, *22*, 294–311.
- 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.
- Ghana Education Service. (2012). *Programme to scale-up quality kindergarten education in Ghana*. Ghana Ministry of Education. Accra: Ghana https://issuu.com/sabretom/docs/10_12_12_final_version_of_narrative_op.
- Ghana Ministry of Education. (2014). *Ghana 2013 national education assessment: Technical report*. Accra, Ghana: MOE.
- Ghana Statistical Service. (2013). *2010 population & housing Census: National analytical report*. Ghana Statistics Service.
- Gwet, K. (2002). Kappa statistic is not satisfactory for assessing the extent of agreement between raters. Retrieved from the Advanced Analytics, LLC website http://advancedanalyticsllc.com/irrhbk/research_papers/kappa_statistic_is_not_satisfactory.pdf.
- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials Quant. Methods Psychol.*, *8*(1), 23–34.
- Hanushek, E. A. (2002). Evidence, politics, and the class size debate. In L. Mishel, & R. Rothstein (Eds.), *The class size debate* (pp. 37–65). Washington, DC: Economic Policy Institute.
- Heyman, R. E., Lorber, M. F., Eddy, J. M., & West, T. V. (2014). Behavioral observation and coding. In H. T. Reis, & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 345–372). New York, NY: Cambridge University Press.
- Institute for Education Science (IES). (2017). *WWC version 4.0 standards handbook*. Washington, D.C.: IES. https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf.
- Jenkins, J. M., Watts, T. W., Magnuson, K., Gershoff, E. T., Clements, D. H., Sarama, J., et al. (2018). Do high-quality kindergarten and first-grade classrooms mitigate preschool fadeout? *J. Res. Educ. Eff.*, 1–36. <https://doi.org/10.1080/19345747.2018.1441347>.
- 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>.
- 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.
- Lewis, L. (2013). *Is there a role for the private sector in education?* Education for global development. Retrieved September 29, 2017 from: <https://blogs.worldbank.org/education/europeandcentralasia/there-role-private-sector-education>.
- Maslach, C., Jackson, S. E., & Leiter, M. P. (1996). *Maslach burnout inventory manual* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.
- McEwan, P. J. (2015). Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments. *Review of Educational Research*, *85*(3), 353–394.
- Moon, B. (2007). School-based teacher development in Sub-Saharan Africa: Building a new research agenda. *Curriculum Journal*, *18*(3), 355–371.
- Murray, D. W., Rabiner, D. L., Kuhn, L., Pan, Y., & Sabet, R. F. (2018). Investigating teacher and student effects of the Incredible Years Classroom Management Program in early elementary school. *Journal of School Psychology*, *67*, 119–133.
- Osei, G. M. (2006). Teachers in Ghana: Issues of training, remuneration and effectiveness. *International Journal of Educational Development*, *26*, 38–51.
- Owusu, G., & Agyei-Mensah, S. (2011). A comparative study of ethnic residential segregation in Ghana's two largest cities, Accra and Kumasi. *Population and Environment*, *32*(4), 332–352.

- Pesando, L. M., Wolf, S., Behrman, J. R., & Tsinigo, E. (2019). Are private kindergartens really better? Examining school choice, parental resources, and children's school readiness in Ghana. *Comparative Education Review*. in press.
- Popova, A., Evans, D. K., Breeding, M. E., & Arancibia, V. (2018). *Teacher professional development around the world: The gap between evidence and practice (World Bank policy research working paper 8572)*. Washington, DC: World Bank Group. <http://documents.worldbank.org/curated/en/349051535637296801/Teacher-Professional-Development-around-the-World-The-Gap-between-Evidence-and-Practice>.
- Pritchett, L. (2013). *The rebirth of education: Schooling ain't learning*. Washington, DC: Center on Global Development.
- 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.
- Republic of Ghana. (2004). *Early childhood care and development policy. Ministry of Women and Children's Affairs: Accra, Ghana*. Retrieved September 11, 2018 from: <http://www.ilo.org/dyn/natlex/docs/ELECTRONIC/88527/101252/F878960202/GHA88527.pdf>.
- Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review*, 94(2), 247–252.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons, Inc.
- Sandefur, J. (2016). *Internationally comparable mathematics scores for fourteen African countries (Working Paper No. 444)*. Washington, DC: Center for Global Development. <https://doi.org/10.2139/ssrn.2893768>.
- 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.
- Seidman, E., & Tseng, V. (2011). Changing social settings: A framework for action. In M. Aber, K. Maton, & E. Seidman (Eds.), *Empowering Settings and Voices for Social Change* (pp. 12–37). New York: Oxford University Press.
- Srivastava, P. (Ed.). (2013). *Low-fee private schooling: Aggravating equity or mitigating disadvantage?* Oxford, UK: Symposium Books.
- UNESCO. (2015). *Education for All global monitoring report regional overview: Sub-Saharan Africa*. Paris: UNESCO.
- UNICEF. (2015). *Ghana's district league table 2015*. http://www.unicef.org/ghana/edia_9697.html.
- United Nations. (2015). *Transforming our world: The 2030 agenda for sustainable development*. New York: Author.
- Vahey, D. C., Aiken, L. H., Sloane, D. M., Clarke, S. P., & Vargas, D. (2004). Nurse burnout and patient satisfaction. *Medical Care*, 42(2 Suppl), 1157–1166.
- Walter, C., & Briggs, J. (2012). *What professional development makes the most difference to teachers (A report sponsored by Oxford University Press)*. Retrieved on . (Accessed 20 July 2015).
- Wamalwa, F. M., & Burns, J. (2018). Private schools and student learning achievements in Kenya. *Economics of Education Review*, 66, 114–124. <https://doi.org/10.1016/j.econedurev.2018.07.004>.
- Welsh, J. A., Bierman, K. L., & Mathis, E. T. (2014). Parenting programs that promote school readiness. In M. Boivin, & K. Bierman (Eds.), *Promoting school readiness and early learning: The implications of developmental research for practice*. NY: Guilford Press.
- Werner, C. D., Linting, M., Vermeer, H. J., & van Ijzendoorn, M. H. (2016). Do intervention programs in child care promote the quality of caregiver-child interactions? A meta-analysis of randomized controlled trials. *Prevention Science*, 17, 259–273.
- Wolf, S. (2018). Impacts of pre-service training and coaching on kindergarten quality and student learning outcomes in Ghana. *Studies in Educational Evaluation*, 59, 112–123. <https://doi.org/10.1016/j.stueduc.2018.05.001>.
- 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., 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>.
- Wolf, S., Turner, L. L., Jukes, M. C., & Dubeck, M. M. (2018). Changing literacy instruction in Kenyan classrooms: Assessing pathways of influence to improved early literacy outcomes in the HALI intervention. *International Journal of Educational Development*, 62, 27–34. <https://doi.org/10.1016/j.ijedudev.2018.02.004>.
- Young, R., & Johnson, D. R. (2015). Handling missing values in longitudinal panel data with multiple imputation. *Journal of Marriage and Family*, 77(1), 277–294.
- Zigarelli, M. A. (1996). An empirical test of conclusions from effective schools research. *The Journal of Educational Research*, 90(2), 103–110.
- Zuilkowski, S. S., Piper, B., Ong'ele, S., & Kiminza, O. (2018). Parents, quality, and school choice: Why parents in Nairobi choose low-cost private schools over public schools in Kenya's free primary education era. *Oxford Review of Education*, 44(2), 258–274.