

Public Disclosure Authorized

Testing and Scaling-up Supply- and Demand-side Interventions to Improve Kindergarten Educational Quality in Ghana

SIEF Midline Report

March 2017 (revised July 2017)

Public Disclosure Authorized

Prepared by:

Sharon Wolf^a, Edward Tsinigo^b, Jere Behrman^c, and J. Lawrence Aber^d

^a Graduate School of Education, University of Pennsylvania

^b Innovations for Poverty Action

^c Department of Sociology and Economics, University of Pennsylvania

^d Global TIES for Children, Institute for Human Development and Social Change, New York University

Table of Contents

List of Tables	iii
List of Figures	iv
Abbreviations	v
Executive Summary	vi
Introduction.....	1
Section 1: Midline Data Collection Process	3
1.1 Overview.....	3
1.2 Questionnaire Designs and Modifications	4
1.2.1 IDELA.....	4
1.2.2 Classroom Environmental Scan	5
1.3.3 Classroom Quality: TIPPS	5
1.3.4 Teacher Survey	5
1.3.5 Caregiver Survey.....	6
1.3.6 School Attendance Records	6
1.3.7 Surveyor evaluation tools.....	6
Section 2. Data Analytic Methods and Results.....	7
2.1 Methods.....	7
2.1.1 Analytic Samples	7
2.1.2 Measures	8
2.2 Analytic Strategy	13
2.3 Results.....	15
2.3.1 Impacts on Teacher Professional Well-being	16
2.3.2 Impacts on Classroom Outcomes.....	16
2.3.3 Impacts on Child Development Outcomes	17
2.3.4 Moderation by Child Characteristics and Public and Private Sector Schools.....	18
2.3.5 Indirect Associations with Child Outcomes through Classroom Quality	19
Section 3: Summary and Discussion of Findings	23
Section 4. Next Steps.....	25
References.....	27
Appendices.....	30

List of Tables

Table 1a. Means and Mean Differences in School, Teacher and Child Characteristics at Baseline, by Treatment Condition	37
Table 1b. Means and Mean Differences in School, Teacher and Child Characteristics at Baseline, by Treatment Condition Relative to control	38
Table 2. Descriptive Statistics and Bivariate Correlations of Outcome Variables at Follow-up	40
Table 3. Intraclass Correlations for Teacher/Classroom and Child Outcomes at Follow-Up	41
Table 4. Impacts on Teacher Professional Well-being and Classroom Quality	42
Table 5. Impacts on Children's School Readiness Outcomes	43
Table 6. Moderation of Treatment Impacts on Children's School Readiness Composite and Individual Domains, by Grade Level	44
Table 7. Moderation of Treatment Impacts on Outcomes, by Public and Private Sector Status	45
Table 8a. Impact Coefficients for Teacher Professional Well-Being Outcomes with TT as the Reference Group	47
Table 8b. Impact Coefficients for Classroom Quality Outcomes with TT as the Reference Group	48
Table 9. Impact Coefficients for Child Outcomes with TT as the Reference Group	49
Table 10. Impact Estimates for Professional Well-Being and Classroom Quality Outcomes Using Only the Stratification Variables as Covariates	50
Table 11. Impact Estimates for Child Outcomes Using Only the Stratification Variables as Covariates	52
Table 12. Attrition from Baseline to Midline for Children and Teachers, by Treatment Condition	54
Table 13. Upper and Lower Bound Estimates for Impact Estimates on Teacher Professional Well-Being	55
Table 14. Upper and Lower Bound Estimates for Impact Estimates on Classroom Quality	56
Table 15. Results of sensitivity analysis for impacts on child outcomes	58
Table 16. Results of sensitivity analysis for impacts on classroom quality	59
Table 17. Impact Estimates with Covariates for Teacher Professional Well-Being Outcomes	61
Table 18. Impact Estimates with Covariates for Classroom Quality Outcomes	63
Table 19. Impact Estimates with Covariates for Child School Readiness Outcomes	65

List of Figures

Figure 1. Research Design of the QP4G Study.....	32
Figure 2. Moderation on Teacher Burnout by Public and Private Schools	53
Figure 3. Predicted Probability of Teacher Turnover by Treatment Status for Public and Private Schools.....	54
Figure 4. QP4G Theory of Change.....	55
Figure 5. Impact Estimates for TT and TTPA Treatment Conditions, by Child Grade Level	56

Abbreviations

ECD	Early Childhood Development
IDELA	International Development and Early Learning Assessment
IPA	Innovations for Poverty Action
KG	Kindergarten
QP4G	Quality Preschool for Ghana Study
TIPPS	Teacher Instructional Practices and Processes System
TT	Teacher Training
TTPA	Teacher Training plus Parent Awareness Training

Executive Summary

The QP4G Study seeks to evaluate the impact of an in-service teacher training and parental awareness training in preschools on teachers, classrooms, and children in Ghana. The midline data collection activities were conducted to obtain information on vital events or changes in the attitudes, perceptions, and behavior of key stakeholders – kindergarten (KG) teachers, children, and caregivers - that had taken place after the baseline visits. Data were collected, using different instruments, to measure three sets of outcomes: (1) child learning and development, (2) classroom quality and teacher well-being, and (3) caregiver outcomes. This report presents results from the impact evaluation of the program after one school-year of intervention.

Using a cluster-randomized control trial, schools were assigned to each of the three treatment arms: teacher training, teacher training plus parent awareness training, and control group; stratified by the district and type of school. Classroom quality and teacher well-being were measured using data on teacher motivation, job satisfaction, burnout, and turnover as well as behavioral management, instructional practice and teacher-child interaction quality. Data on early literacy, early numeracy, socio-emotional development, executive function and approaches to learning among kindergarten children were used as proxies for measuring school readiness. Caregiver outcomes centered on parental perceptions and knowledge of early childhood development and education. Three sets of analytical methods were employed: baseline equivalency, differential attribution, and impact analysis. Impact analysis was done at three levels, namely child level, classroom level, and school level. The implementation and first-year evaluation of the QP4G intervention occurred between September 2015 and June 2016.

The baseline equivalency analysis showed that the results of the randomization process were largely successful, that is, there was not a particular pattern in differences across treatment conditions. We found moderate impacts of the teacher training on reduced burnout and job turnover, improved classroom quality, and improved children’s school readiness. The training program had an impact on supporting student expression and emotional support and behavior management but not on facilitating deeper learning at the classroom level. The school readiness domain with the largest impacts was social-emotional skills, followed by early numeracy. No significant effects were reported for early literacy and executive function. We also found counter-acting effects of the parental awareness training on teachers’ support of student expression in the classroom, as well as child outcomes.

The next line of activities will focus on the analysis of the in-depth qualitative interviews with caregivers and teachers as well as completion of the collection and analysis of the endline data.

Introduction

Recent years have seen a marked increase in both the demand for and the supply of early childhood education services in Ghana. An exploratory study conducted by IPA in 2013 in the Ashaiman neighborhood revealed two key findings: (a) the quality of classroom instruction in preschools was generally low and developmentally inappropriate, and (b) parents' subjective assessment of preschool quality focused on developmentally inappropriate instruction and on classroom materials and infrastructure. Low quality of classroom instructions in preschools in Ghana has mostly been attributed to the fact that most preschool teachers are untrained or inexperienced, as well as a lack of/inadequate in-service training for preschool teachers. In fact, the results of the scoping study revealed that 69% of teachers have no training in education or childhood development. Moreover, even though governmental systems exist to provide feedback to teachers, such systems are rarely used. Parents' subjective assessment of preschool is visible in their evaluation of quality in terms of material infrastructure and perceived "serious lessons" through repetition of letters and numbers. Collectively, low quality of preschool classroom instruction has led to inadequate preparation of children to be ready for progression into the primary school system.

In order to address the above policy concerns, researchers from the University of Pennsylvania and New York University in the United States in partnership with IPA, seek to improve the quality of kindergarten education through teachers and parents. Specifically, the Quality Preschool for Ghana (QP4G) Study involves:

- a. An 8-day in-service teacher training delivered by the National Nursery Teacher Training Center with monitoring and feedback visits;
- b. A 3-part video and discussion intervention delivered to parents through school Parent-Teacher Association meetings focused on early childhood development and learning;
- c. Evaluating the effectiveness of (a) improving the supply of teacher training; and (b) improving the supply and changing the demand of parental intervention.

This report focuses on the second round of data collection conducted for the QP4G project. Midline data collection, or Follow Up 1 (FU1) was conducted in the May-June of 2016 and included surveys with KG teachers and child caregivers; direct assessments of child outcomes; as well as classroom observations of KG teachers. The first round of data collection, baseline data collection, was conducted in the summer and fall of 2015 and included surveys

involving school proprietors, head teachers, KG teachers and child caregivers. This also included the direct assessments of child outcomes; as well as classroom observations of KG teachers. Results from the baseline data collection were discussed in detail in the baseline report submitted in February 2016. This report is organized into four sections. In Section 1, we present a summary of the midline data collection process. In Section 2, we present the data analytic methods and results. In Section 3, we present a summary and discussion of the findings, and in Section 4 we discuss next steps.

Section 1: Midline Data Collection Process

1.1 Overview

Multiple sources of data were collected, including (a) direct assessments of children's school readiness at school entry, (b) surveys of teacher well-being and demographics, (c) video recordings for classroom observations of teachers, and (d) caregiver surveys. The first three occurred at the same time, followed by the caregiver survey. More specifically, the following data were collected:

- a. **Child direct assessments** were conducted on sampled KG children on key indicators of early childhood development (ECD) and administered using the International Development and Early Learning Assessment (IDELA) tool.
- b. **KG teacher survey** sought information on teacher background, poverty status, food security situation at the household level, perceptions about ECD, participation in in-service training, work conditions, teacher well-being, and teaching knowledge.
- c. **KG classroom observation** was done to take inventory of facilities within the KG class; videotape class processes, teaching, and learning; as well as code the video recordings using Teacher Instructional Practices and Processes Systems (TIPPS).
- d. **School attendance records** collected information on school attendance records on sampled KG teachers and children during the 2015/2016 school year.
- e. **Caregiver surveys** were carried out to interview primary caregivers¹ of sampled KG children on their background, poverty status, involvement in school activities, and perception about ECD.

Every attempt was made to ensure that the same children and teachers from baseline were followed. However, due to teacher and child mobility, some teachers and children were “replaced” if they had left the school and could not be located. Consent was obtained for any new participant. Note that for children, passive parental consent was collected for all children in the school at baseline; children whose caregivers asked that they not be included in the study were not selected as replacements.

¹ A primary caregiver is the person who is primarily responsible for a child's care, education and could best talk about his/her educational experiences in school and at home. The primary caregiver may be the child's parent, a family member, guardian, or another individual.

Overall, the completed sample of children at midline was 3,407 with 2,975 of them present both at baseline and midline. The complete sample of teachers at midline was 441, with 347 teachers present at both baseline and midline. Notably, three schools did not wish to participate in the midline assessment. The Head Teacher felt that they were not being compensated properly for the time the assessments took. Also, two schools closed down due to litigation on the school land. Thus, the final sample of schools at midline was 235.

There were minimal issues with midline data collection. The main challenges encountered during the midline data collection activities were mobility and absenteeism of KG teachers and children. There were 81 and 367 reported cases of teacher and child mobility, respectively. Average absenteeism per week during the seven weeks of data collection was 50. Continuous “mop ups” were conducted to track the absentee children; this number then reduced to 40 at the end of the midline data collection period. We could not track these children because they were not available during the period of the data collection.

1.2 Questionnaire Designs and Modifications

The FUP I questionnaires included direct child assessments (measured using the IDELA), direct classroom observations, a teacher survey, school administrative attendance records, and a caregiver phone survey. All these questionnaires, except School Attendance Records and the Interview Protocols, were used at baseline data collection with some minor modifications.

1.2.1 IDELA

The KG Child Assessment was conducted using IDELA tool designed by Save the Children. IDELA was adapted based on extensive pre-testing and piloting by different members of the evaluation team. The adapted version measured five indicators of ECD. The indicators were early numeracy skills, language/literacy skills and development, physical well-being and motor development, socio-emotional development, and approaches to learning. IDELA contained 28 items. In addition, one task was added – the Pencil Tap – to assess executive function skills. IDELA was translated into three local languages, namely, Twi, Ga, and Ewe. These local language versions had gone through rigorous processes of translation and back translation. See *Appendix A Attachment 1* for all materials used for the Child Direct Assessment.

1.2.2 Classroom Environmental Scan

The KG classroom observation involved taking inventories of the KG classrooms [environmental scan] and conducting video recordings of the classroom processes. The KG Class Environmental Scan tool was designed to take inventories of the facilities in the KG classrooms. *Appendix A Attach. 2* gives the data collection and training tools used for the classroom environmental scan and TIPPS.

1.3.3 Classroom Quality: TIPPS

The video recordings taken during the classroom observations were coded using an early childhood education adapted version of TIPPS. Seidman, Raza, Kim, and McCoy (2014) of New York University developed the TIPPS instrument. TIPPS observes nineteen key concepts of teacher practices and classroom processes that influence children's cognitive and social-emotional development. The concept sheet was used to code the kindergarten classroom videos and is provided in *Appendix A Attachment 3*.

1.3.4 Teacher Survey

The FUP I KG Teacher Survey was based on the modification of its baseline version. The modifications took into account the data need for measuring outcomes of the teacher training intervention as well as concerns about respondent burden/distress, response rates, and costs. Two modules were added: (a) participation in in-service training (i.e., participation in any in-service training including QP4G teacher training, issues of contamination or spill-over effects, and receipt of text-message intervention); and (b) perceptions of early childhood development. The latter module was culled from the Caregiver Survey. Also, KG teachers who took part in the Baseline II survey were excluded from answering time-invariant questions. The time-invariant questions/modules were (a) background characteristics such as local languages of teachers/caregivers, the level of proficiency in speaking and writing in English and local languages; teacher's paternal and maternal educational level, and (b) English reading knowledge. Two KG Teacher Audit Surveys were developed from the KG Teacher Survey. The data collection and training materials for the KG Teacher Survey are provided in *Appendix A Attachment 4*.

1.3.5 Caregiver Survey

The FUP I Caregiver Survey was based on the modification of the Baseline II Caregiver Survey. The modification involved the removal of the food security, tracking and mobility updates modules as well as the addition of modules on child discipline and parental participation in the parental awareness raising program. The module on child discipline was adopted from the UNICEF's MICS 2013 Household Questionnaire. A Call Records Form was also designed to track and screen each caregiver before the actual interview was conducted. Four local language dictionaries of keywords and phrases were developed for the Caregiver Survey. The selected languages were Ga, Ewe, Twi, and Hausa. These languages were selected because they were used extensively in interviewing the caregivers at Baseline II survey. Two audit questionnaires were developed for the Caregiver Survey. The Caregiver Survey and training materials can be found in *Appendix A Attachment 5*.

1.3.6 School Attendance Records

The School Attendance Records Form were designed to record school attendance information for the sampled KG teachers and children. The Form captured school-specific attendance details such as the active number of school days, the number of national/school-related holidays per term, and child-specific information such as present/absent from school. The school attendance data collection and training materials are in *Appendix A Attachment 6*.

1.3.7 Surveyor evaluation tools

Surveyor evaluation tools were used to help observe and evaluate the performance of the surveyors during the survey period. The surveyor evaluation tools were Child Assessor Form, Teacher Interviewer Form, and Video Recording Quality Form. The tools were developed to provide a standardized assessment of how the field staff performed when administering the IDELA and KG teacher survey as well as assess the quality of the classroom video recordings. The tools also helped in providing feedback to the surveyors. The Child Assessor and Video Recording Quality Forms were specifically developed for the QP4G Study; tailored towards ensuring strict adherence to the quality protocols prescribed for administering the IDELA and videotaping the KG classroom processes.

Section 2. Data Analytic Methods and Results

2.1 Methods

The implementation and first year evaluation of the Quality Preschool for Ghana intervention occurred between September 2015 and June 2016. The research design is a *cluster randomized controlled trial*, in which schools were assigned to each of the three treatment arms: (1) teacher training (TT; 82 schools), (2) teacher training plus parent awareness training (TTPA; 79 schools), and (3) control group (79 schools). Randomization was stratified by two levels: districts and public and private sector. In addition, stratified by treatment status, treatment schools were then randomly assigned to receive reinforcement messages from the trainings (weekly text messages for teachers ($N = 80$ schools) and picture-based paper flyers for parents ($N = 40$ schools)). The research design is shown in *Appendix B Figure 1*. The trial was registered in the American Economic Associations' registry for randomized controlled trials (RCT ID: AEARCTR-0000704).

2.1.1 Analytic Samples

School sample. All schools in the six districts were identified using the Ghana Education Service Educational Management Information System (EMIS) database, which lists all registered schools in the country. Schools were then randomly sampled stratified by district, and within district by public and private schools. A school listing was then conducted to confirm the presence of each school and to obtain information on each school's head teacher and proprietor. The school listing was also done to obtain information on whether the schools have a kindergarten unit, the inventory of school infrastructure, and GPS coordinates to aid in the randomization of schools. Because there were fewer than 120 public schools across the six districts, every public school was sampled. Private schools were sampled within districts in proportion to the total number of private schools in each district relative to the six districts.

In each district, 20 additional private schools were randomly sampled to serve as "reserve" schools in the event that one of the original schools sampled refused or was not eligible to participate in the study. Examples of needing to use reserve schools were: refusal to participate, discovery that a school does not have a kindergarten program, and that a school listed in the EMIS dataset no longer existed. Eleven schools were replaced from the original 240: six had inaccurate location and contact information and could not be found; three refused to

participate, and two did not have a KG program. The final baseline sample consisted of 240 schools. In the spring follow-up assessment, 2 schools had closed down and 3 schools dropped out of the study (4 control schools, 1 TT school). Thus, the sample for the impact analysis was 235 schools.

Teacher sample. The majority of schools had two KG teachers, though the range was 1–5. All KG teachers in the treatment schools were invited to participate in the training. For the evaluation, 2 teachers (one KG1 and one KG2 teacher) were randomly selected from each school, bringing the total sample at baseline to 444 teachers. Ghana’s education system experiences high rates of teacher mobility and turnover (Osei, 2006). By follow-up 1, 81 teachers were no longer teaching KG at the school. For the impact analysis, we include only teachers who were present at baseline and follow-up (i.e., stayers; Vuchinich et al., 2012). The sample for the impact analysis was 337 teachers.

Child sample. Fifteen children (8 from the KG1 teacher, and 7 from the KG2 teacher) were randomly selected from each class roster to participate in direct assessments. If a school had fewer than 15 kindergarten children enrolled, all children were selected. At baseline, the total sample of children was 3,435 children, with an average of 14.3 children per school (*range* = 4-15). For schools with only one KG classroom, 15 children were randomly sampled from the classroom. For follow-up, we only included children who were still in a study school (including children who transferred to a different school within the study). Total reported mobility of the children was 367, that is, 11%. The final sample for the impact analysis included 2,975 children for whom we have both baseline and follow-up data.

2.1.2 Measures

Teacher professional well-being. Teachers answered an administered survey in English. All items were pilot tested. First, we conducted five cognitive interviews with teachers to assess whether they understood each question, both consistently across constructs and in the way the item was intended (Collins, 2003). We then piloted the survey by administering it to 20 teachers and then assessed the distribution of responses for each item. From both of these exercises, we concluded that these items were suitable for use in this sample. Notably, all items have been used in previous research with teachers in Sub-Saharan Africa (Wolf et al., 2015a; Wolf et al., 2015b). Items were selected from existing scales and factors were derived through exploratory factor analyses conducted with the baseline data. All outcomes were measured at baseline and follow-

up.

Motivation. Teacher's motivation was measured using five items adapted from Bennell & Akyeampong (2007) as reported in Torrente et al. (2012). Items were answered on the following scale: $1 = false$, $2 = mostly\ false$, $3 = sometimes$, $4 = mostly\ true$, $5 = true$. Sample items include "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" ($Mean\ (M) = 4.6$, Standard Deviation (SD) = .59, $\alpha = .77$).

Job Satisfaction. Teacher's job satisfaction was measured using six items adapted from Bennell & Akyeampong (2007) as reported in Torrente et al. (2012). 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 ($M = 3.09$, $SD = .69$, $\alpha = .73$).

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 "feel emotionally drained from my work," "fatigued when I get up in the morning and have to face another day on the job", and "feel burned out from my work." ($M = 2.03$, $SD = .90$, $\alpha = .75$).

Turnover. Teacher turnover ($1 = yes$, $0 = no$) was indicated if the teacher had left his or her position when we returned to the school for follow-up data collection in the third term. If the teacher was absent, we confirmed with the school administration that the teacher had left their position at the school. Approximately one-quarter of teachers from baseline ($N = 107$) had left their position by follow-up.

Classroom outcomes. All teachers were videotaped teaching a lesson in their classrooms for 30-60 minutes in May or June of 2016. Videos were coded with two instruments: an implementation fidelity checklist, and a tool to assess the quality of teacher-child interactions. Both measures are assessed at follow-up only.

Fidelity checklist. We created a checklist of 15 activities that were explicitly covered in the teacher training related to behavior management and instructional practice. 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” ($M = 3.51$, $SD = 2.22$).

Teacher-child interaction quality. All videos were coded using the TIPPS (TIPPS; Seidman et al., 2013; 2017). The TIPPS is a classroom observation tool assessing classroom quality that focuses on the nature of teacher-child interactions created for use in low- and middle-income countries. We used the TIPP-Early Childhood Development 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. (2013) and Seidman et al. (2017).

The TIPPS is made up of 19 items. We drop four items due to lack of variability in their scores across classrooms. Using exploratory and confirmatory factor analysis, we grouped the remaining 15 items into three factors: *Facilitating Deeper Learning* (FDL, 3 items; connecting lesson to teaching objectives; provides specific, high quality feedback; and uses scaffolding; $\alpha = .42$); *Supporting Student Expression* (SSE, 4 items; considers student ideas and interests; encourages students to reason and problem solve; connects lesson to students’ daily lives; and models complex language; $\alpha = .63$), and *Emotional Support and Behavior Management* (ESBM, 7 items; positive climate; negative climate; sensitivity and responsiveness; tone of voice; positive behavior management; provides consistent routines, student engagement in class activities; $\alpha = .83$). See Wolf et al. (2017, under review) for details on the analysis and concurrent validity of the three factors in this sample.

Reliability. Video coders were trained and had to achieve the pre-specified levels of reliability in order to pass the training. Raters were recruited in Ghana, had a bachelors or master’s degree, and attended a five-day training session on the instrument. Each rater had to meet or exceed three-calibration criteria within three attempts to be certified as a TIPPS observer. TIPPS calibration criteria not only looks at agreement but also the degree of deviation

from master codes – both important aspects given that there are only four scale points and that understanding of the concept is critical for precise coding (see Seidman et al. for details on calibration cut-offs). Collectively, these three criteria enhance the likelihood of achieving acceptable levels of inter-rater reliability. Raters that achieved calibration were also required to participate in 30-minute weekly refresher sessions led by TIPPS trainers that included a review of different manual concepts, short practice videos, and time for questions and discussion.

To assess inter-rater reliability, 15% of videos collected at baseline were coded by three raters. We calculated the intraclass correlation coefficient (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. On average across items, 71.1% of the variance was shared across raters.

Child development outcomes. Children’s development was directly assessed in four areas relevant to school readiness: early numeracy, early literacy, social-emotional, and executive function. A fifth domain of children’s approaches to learning was reported by the assessor. The instrument used was the IDELA, developed by Save the Children (Pisani et al., 2015). The tool was translated into three local languages: Twi, Ewe, and Ga. The IDELA was translated, and then back-translated by a different person to check for accuracy. Any discrepancies were discussed and addressed. Finally, after being trained on the instrument, a group of surveyors read and discussed the translated version in their respective local language and made additional changes as a group.

For the main impact analysis, scores on all four domains were combined to create a total “school readiness” score for each child. A previous study has validated the factor structure of the IDELA in Ethiopia as a measure of holistic school readiness (Wolf et al., 2017). Factor analysis was used in the present sample to confirm the use as a 1-factor model of “developmental skills” relevant to school readiness in this sample ($CFI = .975$; $RMSEA = .035$).

Early literacy. The domain of early literacy consists of 38 items grouped into 6 subtasks, and covers constructs of print awareness, letter knowledge, phonological awareness, oral comprehension, emergent writing, and expressive vocabulary. An example subtask on phonological awareness asked children to identify words that begin with the same sound. A sample item is: “Here is my friend mouse. Mouse starts with /m/. What other word starts with /m/? Cow, doll, milk” ($\alpha = 0.74$).

Early numeracy. The domain of early numeracy consists of 39 items grouped into 8

subtasks and covers constructs of number knowledge, basic addition and subtraction, one-to-one correspondence, shape identification, sorting abilities based on color and shape, size and length differentiation, and completion of a simple puzzle. An example item assessing shape identification showed the child a picture with six shapes and asking the child to identify the circle ($\alpha = 0.72$).

Social-emotional development. The domain of social-emotional development consists of 14 items grouped into 5 subtasks, and covers constructs of self-awareness, emotion identification, perspective taking and empathy, friendship, and conflict and problem solving. An example item of conflict solving involved asking the child to imagine he or she is playing with a toy and another child wants to play with the same toy, and asking the child 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, getting another toy ($\alpha = 0.69$).

Executive function. The domain of executive function was assessed with ten items grouped into two subtasks focused on working memory (i.e., forward digit span) and impulse control (i.e., head-toes task). For the forward digit span, assessors read aloud five digit sequences (beginning with two digits and increasing up to six digits) and children were asked to repeat the digit span and marked as correct or incorrect. For the head-toes task, assessors asked children to touch their toes when the assessor touched his or her head, and vice versa in a series of five items ($\alpha = 0.83$).

Approaches to Learning. After the assessor completed the IDELA items with each child, they filled out seven items about the child’s approaches to learning. Each child was rated on a scale of 1 to 4, with 1= “almost never” and 4=“almost always”. Assessors reported on children’s attention (i.e., “Did the child pay attention to the instructions and demonstrations through the assessment?”), confidence, concentration, diligence, pleasure, motivation, and curiosity during the tasks ($\alpha = 0.94$).

Reliability. Inter-rater reliability on the child development outcome measure was assessed. Enumerators were paired and assessed and scored two children together. Cohen’s kappa values were calculated for each pair across each item in the entire assessment, and values ranged from 0.67 to 0.97, with an average kappa value of 0.86.

Covariates. We included a select set of covariates to improve the precision of our impact

estimates. For all models, these included private sector status of the school, six district dummies, a dummy variable for if the school was assigned to receive teacher text messages, a dummy for if the school was assigned to receive parent flyers, and a series of five dummy variables accounting for within-sample mobility (e.g., between baseline and follow-up a baseline school split into two separate schools; two schools merged into one school; children or teachers moved to a different school within the sample). For child outcomes, we also included child gender, age, KG level (1, 2, or 3 if KG1 and KG2 were combined in one classroom, as a categorical variable), and baseline score for each respective outcome. For teacher outcomes, we also included teacher gender, age, level of education, years of teaching experience, and baseline score for each respective outcome.

2.2 Analytic Strategy

Baseline equivalency. We first conducted a baseline equivalency analysis to confirm whether the randomization was successful—i.e., to ensure that the randomization yielded treatment and control groups that are statistically equivalent. We calculated the mean values for a set of school characteristics, teacher characteristics, and child characteristics and baseline school readiness scores by treatment group (see *Appendix B Table 1*). Second, we conducted an omnibus F -test with each set of characteristics in a MANOVA equation with treatment status as the grouping variable, to assess if overall the set of predictors statistically differentiated across treatment groups. If the F -test was not statistically significant, we cannot reject the null hypothesis that the predictors did not differentiate across treatment groups. Overall, there were no meaningful differences across the three treatment arms for baseline school characteristics (Omnibus $F(2) = 0.97, p = 0.520$), teacher characteristics (Omnibus $F(2) = 1.06, p = 0.380$), or child characteristics (Omnibus $F(2) = 0.99, p = 0.429$). Thus, we interpret the few differences between the intervention groups and the control group at baseline as occurring by chance. (See *Appendix C* for a sensitivity analysis).

Differential attrition analysis. Between baseline and follow-up, 81 teachers were no longer working in the school at follow-up, and 367 children transferred or left the school at follow-up. Also, 26 and 93 children and teachers, respectively, were not interviewed at midline because they were unavailable during the entire data collection period or their schools had closed down. We conducted multi-level logistic regression analyses, with an indicator of if the teacher or child left the study sample, to assess if there was differential attrition of teachers or of children by treatment status (internal validity), and other characteristics (external validity). For the

teacher sample, the question of treatment status is considered in the impact analysis since we found significant impacts on teacher turnover. To assess external generalizability of the sample of teachers who stayed, we assessed baseline motivation, burnout and job satisfaction, age, gender, education level, years of teaching experience, and private sector status. Of these ten predictors, only one—baseline job satisfaction—significantly predicted teacher attrition, such that teachers with higher levels of baseline job satisfaction were less likely to leave the study sample ($b = -0.49, SE = 0.25, p < .05$). For the *child sample*, treatment status did not significantly predict whether children left the study sample, indicating that our experimental design was not compromised. To assess external validity of the sample of children that stayed, we assessed baseline levels of school readiness, child gender, child age, and private sector status. We found that baseline school readiness predicted a lower likelihood of leaving the study sample ($b = -0.92, SE = 0.36, p < .05$) and child age predicted a higher likelihood of leaving the study sample ($b = 0.15, SE = 0.05, p < .01$).

Impact analysis. To account for the nested, non-independent nature of the data (i.e., students nested within classrooms and classrooms nested within schools), we employed three-level (for child outcomes) and two-level (for teacher and classroom outcomes) multi-level modeling in Stata (Version 14.0). First, we estimated unconditional models to estimate the intraclass correlations (ICCs), or the proportion of variance in each of the teacher/classroom and student outcomes attributable to students, teachers/classrooms, and schools. Second, impact analyses were conducted with a select set of covariates. We nested children and teachers in the baseline schools from which they were sampled, regardless of their mobility across schools within the sample.

Separate models were fitted to estimate main intervention impacts on (a) teacher professional well-being (i.e., motivation, burnout, and job satisfaction), (b) classroom quality factors (i.e., fidelity checklist, and FDL, SSE, ESBM factors), and (c) children’s school readiness (i.e., total IDELA score). As a post-hoc test, we estimated impacts on each of the four individual domains of children’s school readiness (i.e., early literacy, early numeracy, social-emotional, and executive function) to assess if impacts on child outcomes were driven by any particular domain. The equations for the 3-level model were as follows:

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, baseline score, student mobility dummy variables).

Level 2 (Classroom-level) Model:

$$\mathbf{B}_{0jk} = \gamma_{00k} + \mathbf{u}_{0jk}$$

Where \mathbf{B}_{0jk} is the classroom-level random intercept.

Level 3 (School-level) Model:

$$\gamma_{00k} = \pi_{000} + \pi_{001}\mathbf{T}_k + \pi_{002}'\mathbf{Z}_k + \mathbf{v}_{00k}$$

Where γ_{00k} is the school-level random intercept; \mathbf{Z}_k is the vector of school-level covariates (district dummies, private or public status, within sample mobility dummies); and \mathbf{T}_k is the treatment status assigned to the school.

Third, as a secondary analysis, we examined whether intervention impacts were moderated by child characteristics (gender, child baseline scores, and grade level (KG1 and KG2)) and by school sector (private and public). Moderation of impacts by child covariates was tested by adding a cross-level interaction term between each treatment condition (at level 3) and child characteristic (at level 1). Moderation by sector was calculated with an interaction term (at level 3) between school sector (I =private, O =public) and treatment status. Fourth, we analyzed indirect relationships between treatment status, classroom quality and teacher well-being, and child outcomes using a multilevel structural equation modeling (MSEM) framework as outlined by Preacher, Zyphur and Zhang (2010). This allowed us to descriptively examine (non-causal) mediational relationships between treatment status, the classroom context, and child outcomes.

2.3 Results

The descriptive statistics for all outcome variables, and their intercorrelations, are presented in *Appendix B Table 2*. As a first step, unconditional two-level models were examined for teacher and classroom outcomes at follow-up, and three-level models were examined for child outcomes. The intra-class correlation (ICC), which quantifies the proportion of the total variation for each outcome accounted for by the different levels in the model, was then calculated. Results are shown in *Appendix B Table 3*. The majority of variance in teacher professional well-being and classroom quality (between 74-88%) was accounted for by differences across teachers, rather than across schools. The majority of child outcomes was accounted for by differences across children (between 58-83%) and secondarily across

classrooms (14-42%). A very small portion of the variance (0-9%) was accounted for across schools, indicating that classrooms are more important than schools in explaining variance across child outcomes. This is consistent with variance decomposition studies in the United States (e.g., Rivkin, Hanushek & Kain, 2005).

Impact estimates are presented for the teacher training (TT) condition, and the teacher training plus parent awareness training (TTPA) condition compared to the control condition.

2.3.1 Impacts on Teacher Professional Well-being

Appendix B Table 4 shows the results of analyses estimating the impact of the two treatment conditions on teachers' motivation, burnout, job satisfaction, and teacher turnover. There were no program impacts on either motivation or job satisfaction. The program did impact teacher burnout, reducing burnout in the TT condition ($p < .05$, $d_{wt}^2 = .32$) and the TTPA condition ($p < .001$, $d_{wt} = .51$). Additionally, the TT condition impacted teacher turnover, reducing the probability that a teacher would leave the KG classroom by the third term by 43.5% ($p < .05$, OR = 0.30), reducing turnover from 44.3% of teachers to 26.8%. We found no impacts of the reinforcements to teachers via text message, or the flyers to parents via paper flyers, on any outcome (see Table B1). The coefficients for the full model, including all covariates, are displayed in Appendix Table C1.

2.3.2 Impacts on Classroom Outcomes

Appendix B Table 4 also shows the impact estimates on classroom outcomes. We first addressed the question of fidelity of implementation. We assessed the number of developmentally-appropriate practices observed in the classroom using a checklist of 15 instructional practices that were specifically promoted in the teacher training. The program

² d_{wt} represents a standardized mean difference between treatment and control clusters. 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., $b = .11$), and the three terms of the denominator represent variances at the cluster, school, 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). This same approach was utilized to estimate d_{wt} for this and other main effects presently reported.

increased the number of activities teachers used in the classroom in both treatment conditions by similar magnitudes ($p < .001$, $d_{wt} = .94$). Compared to control group classrooms, who implemented an average of 3.1 activities during the observational assessments, classrooms in both treatment conditions implemented 4.6 activities.

Next, we assessed impacts on classroom quality based on three domains of teacher-child interactions: facilitating deeper learning (FDL; e.g., scaffolding, high quality feedback), supporting student expression (SSE; e.g., considering student ideas during the lesson, encouraging students to reason and problem solve), and emotional support and behavior management (ESBM; e.g., positive climate, teacher sensitivity and responsiveness to student needs, providing consistent routines). There were no impacts of either treatment condition on levels of FDL. Both treatment conditions increased the level of ESBM observed in the classroom ($p < .001$, $d_{wt} = .52$ in the TT condition; $p < .01$, $d_{wt} = .46$ in the TTPA condition). Finally, the TT condition increased levels of SSE in classrooms ($p < .01$, $d_{wt} = .50$), but there were no statistically significant impacts in the TTPA condition.

We found no impacts of the reinforcements to teachers via text message, or the flyers to parents via paper flyers, on any outcome (see Table B2). The coefficients for the full model, including all covariates, are displayed in Appendix Table C2.

2.3.3 Impacts on Child Development Outcomes

Appendix B Table 5 presents the impact estimates of the treatment programs on domains of preschool children's development relevant to school readiness. We first assessed impacts on the composite score of children's developmental readiness for school as our primary outcome of interest. We then conducted post-hoc analyses to assess impacts on each domain of development individually to examine if the findings were driven by any particular developmental domains. The TT program increased children's school readiness ($p < .05$, $d_{wt} = .14$). When broken down by domain, the coefficients for the TT condition were positive for all four domains, but only the social-emotional domain reached statistical significance ($p < .01$, $d_{wt} = .17$; impacts on numeracy were marginal, $p < .057$, $d_{wt} = .09$). Notably, there were no impacts on children's school readiness when the parent training program was added to the teacher training (TTPA).

We found no impacts of the reinforcements to teachers via text message, or the flyers to parents via paper flyers, on any outcome (see Table B3). The coefficients for the full model,

including all covariates, are displayed in Appendix Table C3.

2.3.4 Moderation by Child Characteristics and Public and Private Sector Schools

We assessed moderation in impacts on school readiness by three child characteristics: gender, baseline school readiness, and grade level (KG1 and KG2). We find no statistically significant interactions between treatment status and gender or baseline school readiness (results not shown). We do find statistically significant interactions by grade level for the TTPA condition for the school readiness composite score (see *Appendix B Table 6*). Appendix Figure A1 shows model-adjusted treatment-control group differences for children enrolled in KG1 and KG2 separately. The positive impact of the in-service training on children's school readiness is larger for KG1 than KG2 children.

We then assessed moderation of program impacts on teacher, classroom, and child outcomes by school public versus private sector status. Of the nine primary outcomes assessed, we found two statistically significant interactions between treatment status and public or private sector schools, both in the domain of teacher professional well-being. First, there was a significant interaction between the TT condition and private sector status in predicting teacher burnout ($b = -0.43$, $SE=.21$, $p < .05$). The interaction term was in the same direction and approached statistical significance for the TTPA condition ($b = -0.33$, $SE = .21$, $p = .129$; see *Appendix B Table 7*). *Appendix B Figure 2* illustrates that the impacts on reduced burnout levels were larger in private schools.

Second, the interaction terms predicting teacher turnover between private school status and both the TT and TTPA conditions were marginally statistically significant ($b = -1.32$, $SE=.77$, $p < .10$ and $b = -1.38$, $SE=.78$, $p < .10$, respectively). *Appendix B Figure 3* illustrates the nature of these differences, showing predicted probability of teacher turnover by treatment condition in private and public sector schools separately. The treatment reduced the predicted probability of teacher turnover from 43.5% to 12.3% (TT condition) and 17.4% (TTPA condition). Notably, in private schools the treatment reduced turnover to levels similar to the public sector.

2.3.5 Indirect Associations with Child Outcomes through Classroom Quality

Guided by our theory of change (see *Appendix B Figure 4*), we tested for indirect (mediational) associations between classroom quality, teacher burnout, and children's total school readiness for treatment. We examined the three potential mediators for which we observed positive program impacts – teacher burnout, classroom levels of supporting student expression, and classroom levels of emotional support and behavior management. Because of the mixed findings in the TTPA treatment condition arm, we examine whether there were any indirect associations in the TT condition arm only. We follow the approach recommended by Preacher, Zyphur, and Zhang (2010) and include all mediators in one analysis. We conducted this analysis in MPlus version 6.12. We found evidence of significant direct effects ($b = .024$, S.E. = .011, $p = .022$) and significant indirect effects ($b = .007$, S.E. = .003, $p = .051$). There were significant indirect effects for emotional support and behavior management (ESBM; $b = .006$, S.E. = .003, $p = .023$), which accounted for 25.0% of the total effects. While not statistically significant, teacher burnout accounted for 8.3% of the total effects ($b = .002$, S.E. = .002, $p = .203$), and supporting student expression (SSE) accounted for 0.0%. These descriptive analyses suggested that the positive impact of in-service teacher training on children's development of school readiness skills is partially mediated by the impact on teachers' use of emotional support and positive behavior management strategies.

2.3.6 Treatment Condition Contrasts

We next ran main impact models estimated with the TT (rather than control condition) as the reference group. Table 8 presents impact estimates on teacher professional well-being and classroom quality, and Table 9 presents impact estimates for child outcomes, providing the direct contrast between TT and TTPA (in addition to TT vs. Control and TTPA vs. control). Regarding teacher professional well-being outcomes, while coefficient sizes differ across TT and TTPA conditions, none of these differences are statistically significant. Regarding classroom quality outcomes, the results show no differential impacts on FDL and ESBM dimensions of quality. The impacts on SSE, however, are statistically different, with TTPA showing smaller program impacts than TT ($b = -.23$, $p < .01$). Notably, as shown in the original models, levels of classroom quality in the TTPA condition are not statistically different than the control group.

Finally, the results show that the counteracting effects of the TTPA condition result in statistically significant differences between TT and TTPA regarding children's school readiness ($b = -.020, p < .01$), specifically in the areas of academic skills including early numeracy ($b = -.028, p < .001$) and early literacy ($b = -.024, p < .05$).

2.3.7 Lower- and Upper- Bound Estimates

As reported above, both treatment arms reduced teacher turnover. Because of differential attrition rates in the treatment and control groups, we ran a sensitivity analysis for teacher-level outcomes. Table 12 shows the attrition for children and teachers from baseline to midline broken down by treatment condition. Notably, there is no differential attrition for children across treatment conditions (TT: $b = -.002, SE = .247, p = .993$, and TTPA: $b = -.076, SE = .248, p = .760$). We now describe our procedure and results to obtain lower- and upper-bound estimates of our treatment impact estimates given teacher attrition. The basic idea of bounding estimates is to see how much the estimates change under extreme assumptions about the values for the missing (attrited) observations (e.g., Lee, 2009).

To obtain *lower-bound estimates*, we used two approaches. First, we follow an approach used by Behrman, Parker, Todd & Wolpin (2015). Using an extensive set of baseline teacher characteristics (including motivation, burnout, job satisfaction, age, education level, gender, years of teaching experience, district, public vs private school, training in ECD, mental health, food security, temporary vs. permanent position, and perceptions of parent support), we used propensity score modeling to match treatment teachers with a paired control group that was at both baseline and follow-up waves. For treatment teachers who attrited, we then imputed their midline score using the score of their matched-pair control group counterpart. Second, rather than match treatment teachers, we assume that all treatment teachers who attrited would have performed at the 25th percentile score of the treatment teachers who stayed. Thus, for teachers who attrited, we imputed their score to be that of the 25th percentile of treatment teachers in their respective treatment condition who were in the sample at midline.

To obtain upper-bound estimates, we assume that all teachers who attrited would have performed at the 75th percentile score of the teachers who stayed. Thus, for teachers who attrited, we imputed their score to be that of the 75th percentile of treatment teachers in their respective treatment condition who were in the sample at midline.

Overall, the pattern of results is the same across the bounded estimates. In other words, impacts that were statistically significant in the original models continue to be significant in both lower- and upper-bound estimates. And, impacts that were not statistically significant in the original models are also not statistically significant in both lower- and upper-bound estimates.

Table 13 displays the results of the lower- and upper- bound estimates for teacher professional well-being outcomes. All estimates followed the same pattern of statistical significance as the main impact estimates, with the two exceptions of impacts on motivation and job satisfaction for teachers in the TT condition reaching marginal statistical significance in the upper-bound estimates. Specifically, for motivation, impact estimates ranged from $b = .057, p = .349$ to $b = .121, p = .054$ for TT, and $b = .006, p = .926$ to $b = .048, p = .447$ for TTPA. No bounded estimate was statistically different from zero. Note that some of the lower-bound estimates are in fact *higher* than the original impact estimates though not significantly different.

For burnout, impacts were all statistically significant in both lower- and upper-bound estimates. For the TT condition, impact estimates ranged from $b = -.237, p = .032, d_{wt} = .284$ to $b = -.412, p < .001, d_{wt} = .476$ (main impact estimate $d_{wt} = .321, p < .05$). For the TTPA condition, impact estimates ranged from $b = -.315, p = .014, d_{wt} = .354$ to $b = -.551, p < .001, d_{wt} = .637$ (main impact estimate $d_{wt} = .507, p < .001$).

For job satisfaction, impact estimates ranged from $b = .012, p = .898$ to $b = .174, p = .074$ for TT, and $b = -.085, p = .375$ to $b = .133, p = .183$ for TTPA.

Table 14 displays the results of the lower- and upper-bound estimates for the three observed classroom quality factors. All estimates followed the same pattern of statistical significance as the main impact estimates. For FDL, impact estimates ranged from $b = -$

.005, $p = .880$ to $b = .094$, $p = .356$ for TT, and $b = -.138$, $p = .196$ to $b = -.018$, $p = .859$ for TTPA. No estimate was statistically different from zero.

For ESBM, impacts were all statistically significant in both lower- and upper-bound estimates. For the TT condition, impact estimates ranged from $b = .172$, $p < .002$, $d_{wt} = .458$ to $b = .229$, $p < .001$, $d_{wt} = .621$ (main impact estimate $d_{wt} = .520$, $p < .001$). For the TTPA condition, impact estimates ranged from $b = .103$, $p = .073$, $d_{wt} = .274$ to $b = .205$, $p < .001$, $d_{wt} = .556$ (main impact estimate $d_{wt} = .459$, $p < .01$).

Finally, for SSE, for the TT condition impact estimates ranged from $b = .245$, $p = .007$, $d_{wt} = .378$ to $b = .388$, $p < .001$, $d_{wt} = .589$ (main impact estimate $d_{wt} = .499$, $p < .01$). For the TTPA condition, no impact estimate was statistically different from zero, similar to the main impact analysis. Impact estimates ranged from $b = .015$, $p = .881$ to $b = .160$, $p = .117$.

Section 3: Summary and Discussion of Findings

This report presented results from an impact evaluation of in-service teacher training and parental awareness training in preschools on teachers, classrooms and children in Ghana after one school-year of intervention. We found moderate impacts of the teacher training on some dimensions of teacher professional well-being (reduced burnout and job turnover), improved classroom quality, and improved children's school readiness. Two domains of classroom quality were impacted – supporting student expression and emotional support and behavior management – but not the third, facilitating deeper learning. Post hoc analyses indicated that the school readiness domain most impacted was social emotional skills. There were marginal impacts on early numeracy, but no significant effects on early literacy or executive function.

The QP4G training included didactic trainings before the school-year started and in-classroom coaching and mentoring over the course of the school year, all implemented by local professionals (including teacher trainers and district government education coordinators). The in-service training and coaching helped teachers incorporate play-based and child-centered methods into literacy and numeracy lessons, as well as develop behavior and classroom management skills. The trainings did not focus on instructional pedagogy regarding language, literacy, and math skills. Thus, it is not surprising that the impacts were observed on social, emotional, and behavior management aspects of the classroom environment (rather than pedagogical).

Research in Chile found similar effects of in-service teacher training on observed levels of classroom emotional support but not instructional support (Yoshikawa et al., 2015), concluding that a focus on behavior management, along with teachers' perceptions that they were receiving support, may have led to increased warm and respectful interactions and positive emotions and expectations in the classroom. Notably, while their study found improvements in classroom quality, these did not translate to improved child outcomes. Similarly, Ozler et al. (2016) found that an intensive, 5-week teacher training in Malawi child-care centers improved classroom quality, but not child outcomes. Thus, it is notable that the less intensive and less costly training evaluated in this study improved both classroom quality and children's outcomes. Given the focus of teacher training, it is not surprising that the impact of the intervention on children's school readiness was primarily on the social-emotional domain. Future data collection will provide evidence on whether these impacts are sustained into the next academic year for both teachers and children, and whether the impacts observed on children's social-emotional

outcomes will translate over time into impacts in academic domains as well (e.g., Graziano, Reaves, & Calkins, 2007; McClelland, Morrison & Holmes, 2000).

The effect sizes we observed ($d=.33$ to $.50$ for teacher and classroom measures; $d=.14$ to $.17$ for child measures) are consistent with the small- to moderate-size effects of similar successful programs in other LMICs (McEwan, 2015; Ozler et al., 2015; Yoshikawa et al., 2015), and with other ECE interventions in the United States (e.g., Morris et al., 2014; Raver et al., 2008). This suggests that future initiatives should focus on how to achieve larger impacts if early education strategies are to have the dramatic effects on children's learning trajectories required to help all children learn.

Our findings are promising from an intergenerational poverty reduction perspective and somewhat promising from an equity perspective. The targeted children were from relatively disadvantaged districts in the Greater Accra Region of Ghana, and the results suggested that the targeted children gained modestly in absolute terms, which should reduce the probabilities that these children live in poverty as adults, and relative to relatively better-off children, thereby reducing overall inequality. The equally positive effects for boys and girls and for the relatively more and less school-ready children also suggests the program did not increase inequalities among the targeted relatively disadvantaged population, in contrast to some recent results for schooling programs in another developing country such as Bangladesh (Behrman, 2015). But also, these equally positive effects for boys and girls and for the relatively more and less ready children means that the program did not reduce inequalities among the targeted relatively poor population.

Our findings also suggested that there are significant gains from teacher training in both private and public schools, and very large reductions in teacher burnout and turnover in the private sector. If the teacher training is publicly provided, then both private and public schools and teachers are likely to have incentives to accept this training. If schools or teachers have to pay for the training, then whether they accept the training presumably depends on whether the perceived gains outweigh the costs. Given the reductions in teacher turnover in the private sector, private schools may find it worthwhile to pay for teachers to attend such training. From the point of view of improving the education of children, however, it is not clear why society would want to create differential incentives for teacher training depending on the ownership of the schools.

Contrary to our prediction, we found that adding a parent awareness training,

administered through school PTA's by local government district coordinators, did not improve the effectiveness of the teacher training. Rather, we found the parent awareness training counteracted some of the positive impacts of the teacher training, specifically the improvements in classroom emotional support and behavior management, and on children's school readiness outcomes. Importantly, the counteracting effects of adding parent awareness training to the teacher training were observed for the older (KG2) children only. Perhaps parents see these messages as more relevant to younger children (4 year olds), and may see older children (5-6 year olds) as needing to get ready for primary school. Or perhaps it was something about the content of the parent awareness training itself? The parent trainings consisted of screened, staged videos in the local language of two mothers discussing the preschool education of their children, and featured the two different classrooms and teachers that were being discussed. It is possible that these videos did not relate to caregivers' experiences, and as a result it caused them to distance themselves from the schools and their child's education. Alternatively, it is possible that the trainings were not implemented with fidelity and that parents' experiences varied widely based on the district education coordinator that was implementing the program. Anecdotally, this appears to be the case. Thus, our conclusion is not that all types of parent awareness training is necessarily harmful to children, but rather that it must be done carefully and in a way that successfully reaches parents. Notably, a recent study in Malawi found that a more intensive, 12-module group-based parenting support program administered through child-care centers by teachers and their mentors combined with intensive teacher training was effective in improving early childhood developmental outcomes (Ozler et al., 2016), suggesting that parenting programs administered through schools by local personnel can be effective. However, it is possible that such programs need to meet frequently enough for parents to internalize the messages.

Finally, we tested the added value of providing teachers and parents with reinforcements of the messages of the trainings via bi-weekly text messages for teachers and paper, picture-based flyers delivered to parents three times in the second and third terms of the academic year. We found no consistent impacts of these additional "nudge-like" reinforcements.

Section 4. Next Steps

Three primary next steps are currently underway. First, as a follow-up to the unexpected findings on the counter-acting effects of the parental awareness training, we have conducted

interviews with 25 caregivers and 25 teachers who were in this treatment arm. These interviews will allow us to gain insight into the experiences of caregivers and teachers regarding this program and may shed light on some of the underlying processes that gave way to the findings from the quantitative data. These interviews are currently being analyzed and results should be ready by late April.

Second, we are currently in the field for endline data collection, nine months after midline data collection occurred. Children who were in KG1 during the implementation year will likely be in KG2, and children who were in KG2 during implementation year will likely be in Primary 1. We have collected the same types of data as midline. While initially we only had enough funding to follow two-thirds of the children at endline, and no teachers, we have secured funding from the Early Learning Partnership at the World Bank to follow the full sample of children, teachers, and caregivers to assess if impacts were sustained the year following implementation.

Third, we are preparing several academic manuscripts based on the data collected to date. Four manuscripts using data from this project are currently under review at academic journals, including a paper summarizing the impacts of the implementation year of the study. The references are as follows:

1. Wolf, S., & McCoy, D. C. (in press). Household Socioeconomic Status and Parental Investments: Direct and Indirect Relations with School Readiness in Ghana. *Child Development*.
2. Wolf, S., Aber, J.L., & Behrman, L. Experimental Evaluation of the ‘Quality Preschool for Ghana’ Intervention on Teacher Professional Well-Being, Classroom Quality and Children’s School Readiness. *Child Development*.
3. Wolf, S., Raza, M., Kim, S., Aber, J.L., Behrman, J., & Seidman, E. (revise and resubmit). Measuring classroom process quality in pre-primary classrooms in Ghana using the TIPPS. *Early Childhood Research Quarterly*.
4. Chan, W. The Relation Between Power in Normal and Binomial Outcomes in Cluster Randomized Trials. *Journal of Research in Educational Effectiveness*.

References

- Behrman, J. A. (2015). Do targeted stipend programs reduce gender and socioeconomic inequalities in schooling attainment? Insights from rural Bangladesh. *Demography*, 52(6), 1917.
- Behrman, J. R., Parker, S. W., Todd, P. E., & Wolpin, K. I. (2015). Aligning learning incentives of students and teachers: Results from a social experiment in Mexican high schools. *Journal of Political Economy*, 123(2), 325-364.
- Bennell, P., & Akyeampong, K. (2007). *Teacher Motivation in Sub-Saharan Africa and South Asia*. London: DfID.
- Collins, D. (2003). Pretesting survey instruments: an overview of cognitive methods. *Quality of Life Research*, 12(3), 229-238.
- Graziano, P. A., Reavis, R. D., Keane, S. P., & Calkins, S. D. (2007). The role of emotion regulation in children's early academic success. *Journal of school psychology*, 45(1), 3-19.
- Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, 76(3), 1071-1102.
- Maslach, C., Jackson, S. E. & Leiter, M. P. (1996). *Maslach Burnout Inventory manual*. 3rd ed. Palo Alto, CA: Consulting Psychologists Press.
- McClelland, M. M., Morrison, F. J., & Holmes, D. L. (2000). Children at risk for early academic problems: The role of learning-related social skills. *Early childhood research quarterly*, 15(3), 307-329.
- 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.
- Morris, P., Mattera, S. K., Castells, N., Bangser, M., Bierman, K., & Raver, C. (2014). *Impact Findings from the Head Start CARES Demonstration: National Evaluation of Three Approaches to Improving Preschoolers' Social and Emotional Competence*. OPRE Report 2014-44. MDRC.
- Osei, G.M. (2006). Teachers in Ghana: Issues of training, remuneration and effectiveness.

International Journal of Educational Development, 26, 38–51.

- Özler, B. , Fernald, L.C.H., Kariger, P., McConnell, C., Neuman, M., & Fraga, E. (2016). *Combining Preschool Teacher Training with Parenting Education: A Cluster-Randomized Controlled Trial*. Policy Research Working Paper 7817. Washington, DC: World Bank.
- Pisani, L., Borisova, I., & Dowd, A. J. (2015). International Development and Early Learning Assessment Technical Working Paper. *Save the Children*. Washington, DC.
- Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, 15(3), 209.
- Raver, C. C., Jones, S. M., Li-Grining, C. P., Metzger, M., Champion, K. M., & Sardin, L. (2008). Improving preschool classroom processes: Preliminary findings from a randomized trial implemented in Head Start settings. *Early Childhood Research Quarterly*, 23(1), 10-26.
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417-458.
- Seidman, E., Raza, M., Kim, S., & McCoy, J. M. (2013). Teacher Instructional Practices and Processes System (V.5) – TIPPS: Manual and scoring system. New York University.
- Seidman, E., Raza, M., & Kim, S. (under review). Assessment of pedagogical practices and processes in low & middle income country secondary school classrooms: Findings from Uganda.
- Torrente, C., Aber, J.L., Witteveen, D., Gupta, T., Johnston, B., Shivshanker, A., Annan, J., & Bundervoet, T. (2012). Baseline Report: Teacher Survey Results. New York University.
- Vuchinich, S., Flay, B. R., Aber, L., & Bickman, L. (2012). Person mobility in the design and analysis of cluster-randomized cohort prevention trials. *Prevention Science*, 13(3), 300-313.
- Wolf, S., Torrente, C., Frisoli, P., Weisenhorn, N., Shivshaker, A., Annan, J., & Aber, J. L. (2015a). Preliminary impacts of a professional development intervention on teacher wellbeing in the Democratic Republic of the Congo. *Teaching and Teacher Education*, 51, 24-36.
- Wolf, S., Aber, J. L., Torrente, C., McCoy, M. & Rasheed, D. (2015b). Cumulative risk and teacher

well-being in the Democratic Republic of the Congo. *Comparative Education Review*, 59(4),717.

Wolf, S., Halpin, P.F., Yoshikawa, H., Down, A., Pisani, L., & Borisova, I. (2017, under revision).

Measuring school readiness globally: Assessing the construct validity and measurement invariance of the IDELA in Ethiopia.

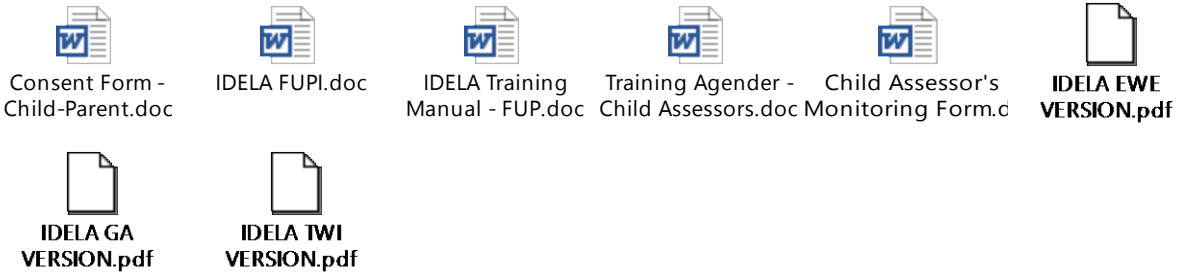
Yoshikawa, H., Leyva, D., Snow, C. E., Treviño, E., Barata, M., Weiland, C., ... & Arbour, M.C.

(2015). Experimental impacts of a teacher professional development program in Chile on preschool classroom quality and child outcomes. *Developmental Psychology*, 51(3), 309.

Appendices

Appendix A

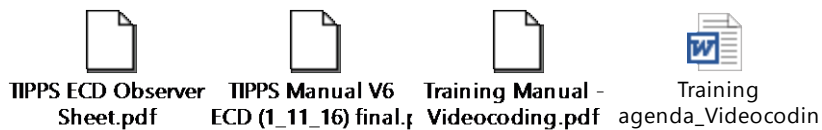
Attachment 1



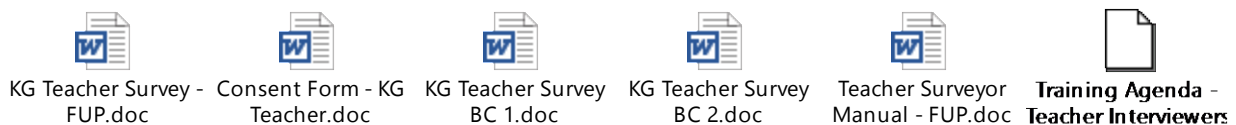
Attachment 2



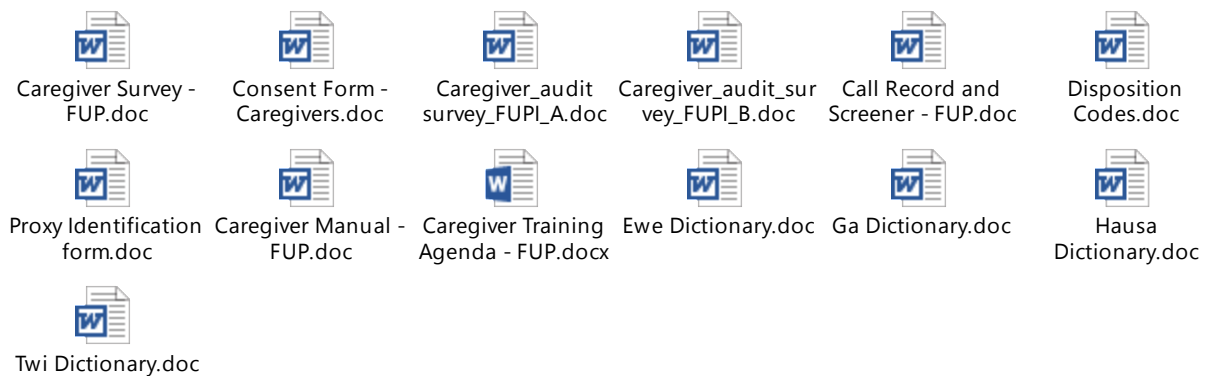
Attachment 3



Attachment 4



Attachment 5



Attachment 6



School Attendance Record - Guidelines.School Attendance Record - Children.doc School Attendance Record - Teachers.doc

Appendix B

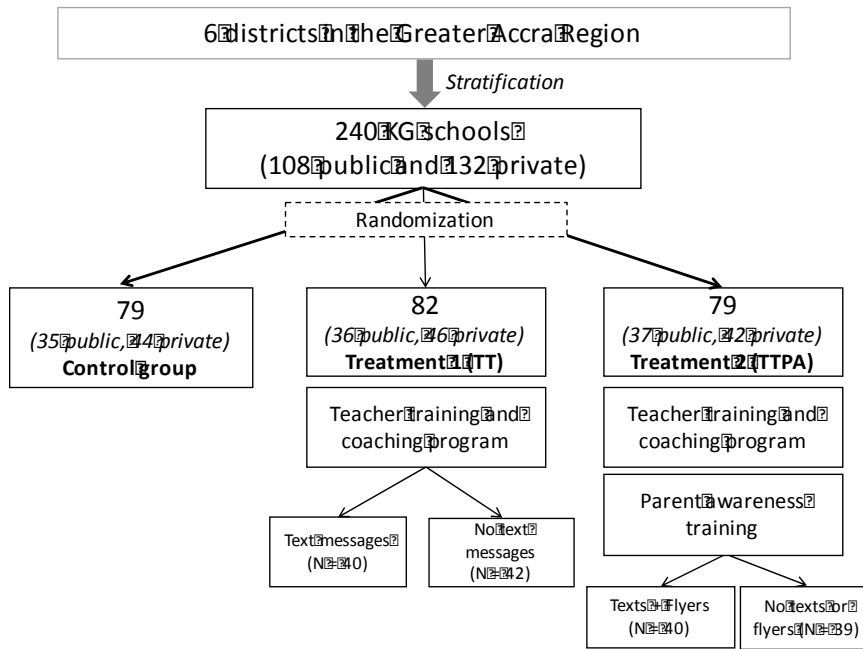


Figure 1. Research Design of the QP4G Study

Note. The six districts are: Ga South, Adenta, Ledzokuku-Krowor, Ga Central, La Nkwantanang-Madina, and Ga West.

Table 8a. Means and Mean Differences in School, Teacher and Child Characteristics at Baseline, by Treatment Condition

	Control	TT	TTPA	<i>F</i> -statistic	<i>p</i> -value
<u>School characteristics</u>					
	Mean or %				
Private school status	55.7%	56.1%	53.2%	0.08	0.923
No. of years school has been established	23	23	19	0.95	0.389
School has written rules/regulations for staff	38.5%	48.8%	35.9%	1.52	0.222
Total number of KG children in school	54	63	60	0.64	0.529
Total number of KG teachers on the payroll	2.0	2.3	2.2	0.98	0.376
Main language of instruction in KG1					
English only	10.5%	13.5%	7.5%	0.68	0.509
Mother tongue only	4.5%	1.4%	1.5%	0.90	0.407
Mixture of English and Mother tongue	85.1%	85.1%	91.0%	0.70	0.496
<u>Head teacher characteristics</u>					
Head teacher has training in ECD	41.0%	48.8%	44.9%	0.11	0.895
Years of experience of head teacher	5.8	6.2	4.9	3.24	0.041
Satisfied with job at school (very true)	70.5%	71.3%	67.9%	0.11	0.900
Wants to transfer to another school (false)	80.8%	83.8%	82.1%	0.12	0.888
Wants to leave education profession (false)	71.8%	86.3%	83.3%	2.96	0.054
Sample size (total = 240)	79	82	79		
<u>Teacher characteristics</u>					
Female	97.9%	97.4%	97.3%	0.05	0.953
Age	35.3	35.7	35.2	0.07	0.933
Father's education level (at least SSS/SHS)	45.0%	53.3%	38.7%	3.30	0.038
Mother's education level (at least SSS/SHS)	30.7%	32.5%	18.0%	4.81	0.009
Years as a teacher	6.55	6.16	6.64	0.22	0.801
Years as a teacher in current school	3.37	3.47	3.21	0.17	0.842
At least secondary high school (%)	97.1%	93.5%	91.3%	2.18	0.114
Has any post-secondary training	60.0%	62.3%	58.7%	0.22	0.804
Has training in ECD	65.7%	72.1%	64.0%	1.25	0.288
Sample size (total = 444)	140	154	150		
<u>Child characteristics</u>					
Female	50.0%	48.5%	49.0%	0.27	0.764
Age	5.25	5.17	5.25	1.02	0.361
KG1 (vs. KG2)	53.5%	52.1%	52.6%	0.24	0.789
School readiness composite (% correct)	50.9	51.8	52.2	1.66	0.19
Early literacy	43.6	44.6	46	3.80	0.023
Early numeracy	38.8	39	40	1.54	0.214
Social-emotional	36.3	37.2	38.4	3.61	0.027
Executive function	46.4	45.9	46.3	0.21	0.814
Sample size (total = 3,435)	1088	1180	1167		

Table 1b. Means and Mean Differences in School, Teacher and Child Characteristics at Baseline, by Treatment Condition Relative to control

	Control	TT	TTPA	TT vs. C		TTPA vs. C	
				t-stat	p-value	t-stat	p-value
<u>School characteristics</u>							
	Mean or %						
Private school status	55.70%	56.10%	53.20%	-0.05	0.959	0.32	0.751
No. of years school has been established	23	23	19	0.11	0.914	1.20	0.231
School has written rules/regulations for staff	38.50%	48.80%	35.90%	-1.30	0.195	0.33	0.742
Total number of KG children in school	54	63	60	-1.01	0.312	-0.98	0.328
Total number of KG teachers on the payroll	2	2.3	2.2	-1.40	0.164	1.03	0.303
Main language of instruction in KG1							
English only	10.50%	13.50%	7.50%	-0.56	0.580	0.60	0.549
Mother tongue only	4.50%	1.40%	1.50%	1.11	0.267	1.01	0.314
Mixture of English and Mother tongue	85.10%	85.10%	91.00%	-0.49	0.625	0.35	0.729
<u>Head teacher characteristics</u>							
Head teacher has training in ECD	41.00%	48.80%	44.90%	-0.97	0.332	-0.48	0.630
Years of experience of head teacher	5.8	6.2	4.9	-0.54	0.590	2.22	0.028
Satisfied with job at school (% very true)	70.50%	71.30%	67.90%	-0.10	0.919	0.35	0.731
Satisfied with decision to be head teacher (% very true)	84.60%	88.80%	94.90%	-0.76	0.919	-2.13	0.035
Wants to transfer to another school (% false)	80.80%	83.80%	82.10%	-0.49	0.626	-0.21	0.838
Wants to leave the education profession (% false)	71.80%	86.30%	83.30%	-2.26	0.025	-1.73	0.085
Sample size (total = 240)	79	82	79				
<u>Teacher characteristics</u>							
Female	97.90%	97.40%	97.30%	0.26	0.799	0.30	0.772
Age	35.3	35.7	35.2	-0.11	0.909	0.24	0.810
Father's education level (at least SSS/SHS)	45.00%	53.30%	38.70%	-1.41	0.159	1.09	0.276
Mother's education level (at least SSS/SHS)	30.70%	32.50%	18.00%	-0.32	0.748	2.55	0.011
Years as a teacher	6.55	6.16	6.64	0.50	0.616	-0.12	0.906
Years as a teacher in current school	3.37	3.47	3.21	-0.21	0.837	0.36	0.721
At least secondary high school (%)	97.10%	93.50%	91.30%	1.46	0.145	2.11	0.035
Has any post-secondary training	60.00%	62.30%	58.70%	-0.41	0.682	0.23	0.818
Has training in ECD	65.70%	72.10%	64.00%	-1.18	0.240	0.31	0.761
Sample size (total = 444)	140	154	150				
<u>Child characteristics</u>							
Female	50.00%	48.50%	49.00%	0.73	0.468	0.47	0.640
Age	5.25	5.17	5.25	1.20	0.232	0.01	0.989
KG1 (vs. KG2)	53.50%	52.10%	52.60%	0.07	0.945	1.12	0.264

Total school readiness (% correct)	50.9	51.8	52.2	-1.10	0.270	-1.92	0.055
Early literacy	43.6	44.6	46	-1.21	0.228	-2.76	0.006
Early numeracy	38.8	39	40	-0.30	0.763	-1.64	0.101
Social-emotional	36.3	37.2	38.4	-1.11	0.267	-2.70	0.007
Executive function	46.4	45.9	46.3	0.61	0.542	0.14	0.891
Sample size (total = 3,435)	1088	1180	1167				

Table 9. Descriptive Statistics and Bivariate Correlations of Outcome Variables at Follow-up

	Mean	SD	Range	1	2	3	4	5	6
1 Child school readiness composite	0.565	0.178	0 - 1						
2 Teacher motivation	4.71	0.44	1 - 5	0.035					
3 Teacher burnout	2.01	0.90	1 - 6	-0.040	-0.174				
4 Teacher job satisfaction	3.08	0.68	1 - 4	-0.017	0.130	-0.284			
<u>Observed classroom quality</u>									
5 Facilitating Deeper Learning	2.39	0.65	1 - 4	0.042	-0.076	-0.051	0.031		
6 Supporting Student Expression	3.07	0.37	1 - 4	0.104	-0.052	-0.080	0.063	0.304	
7 Emotional Support & Behavior Management	1.75	0.67	1 - 4	-0.035	-0.052	0.014	0.000	0.360	0.157

Notes. Bold numbers indicate correlation is statistically significant at $p < .05$. Correlations with school readiness use child-level data ($N = 2,975$); correlations among teacher variables include teacher-level data ($N = 337$).

Table 10. Intraclass Correlations for Teacher/Classroom and Child Outcomes at Follow-Up

	Proportion of Variance		
	Child	Teacher / Classroom	School
<u>Teacher professional well-being</u>			
Motivation		0.874	0.126
Burnout		0.883	0.117
Job satisfaction		0.731	0.269
<u>Classroom quality</u>			
FDL		0.778	0.222
ESBM		0.867	0.133
SSE		0.852	0.148
<u>Child outcomes</u>			
School readiness composite	0.581	0.419	0.000
Early numeracy	0.622	0.378	0.000
Early literacy	0.543	0.371	0.086
Social-emotional	0.826	0.141	0.034
Executive function	0.765	0.199	0.037
Approaches to learning	0.689	0.085	0.227

Notes. FDL = Facilitating Deeper Learning; SSE = Supporting Student Expression; ESBM = Emotional Support and Behavior Management.

Table 11. Impacts on Teacher Professional Well-being and Classroom Quality

	b	SE	p-value		effect size (<i>d_{wt}</i>)
<u>Teacher professional well-being</u>					
Motivation					
TT	0.078	0.068	0.256		0.180
TTPA	-0.031	0.071	0.660		-0.071
Burnout					
TT	-0.286	0.125	0.022	*	0.321
TTPA	-0.452	0.130	0.000	***	0.507
Job Satisfaction					
TT	0.089	0.100	0.375		0.131
TTPA	0.000	0.100	0.999		0.000
Teacher turnover ^a					
TT	-1.203	0.487	0.013	*	
TTPA	-0.745	0.458	0.104		
<u>Classroom outcomes</u>					
Fidelity checklist (# of activities)					
TT	1.495	0.258	0.000	***	0.937
TTPA	1.494	0.265	0.000	***	0.936
Facilitating Deeper Learning (FDL)					
TT	0.016	0.107	0.880		0.025
TTPA	-0.052	0.109	0.663		-0.081
Emotional Support & Behavior Management (ESBM)					
TT	0.196	0.057	0.001	***	0.520
TTPA	0.173	0.059	0.003	**	0.459
Supporting Student Expression (SSE)					
TT	0.321	0.106	0.002	**	0.499
TTPA	0.092	0.109	0.398		0.143
Sample size = 337 teachers/classrooms					

Notes. Estimates are computed using observed scores, in two level models: teachers nested in schools. Effect sizes calculated accounting for the 2-level model structure (Hedges, 2009).

Sample includes teachers present at baseline and follow-up.

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

TT = Teacher training condition; TTPA = Teacher training plus parent awareness training condition.

^a Impacts on turnover (a binary variable) were assessed using a multi-level logistic regression and included the full sample of teachers from baseline ($N = 444$).

Models include the following control variables: private (vs. public) sector status of the school, six district dummies, a dummy variable for if the school was assigned to receive teacher text messages, a dummy for if the school was assigned to receive parent flyers, a series of five dummy variables accounting for within-sample mobility, teacher gender, age, level of education, years of teaching experience. Models for teacher professional well-being outcomes also include the baseline score for each respective outcome.

Table 12. Impacts on Children's School Readiness Outcomes

	b	SE	p-value		<i>effect size (d_w)</i>
<u>School readiness composite</u>					
TT	0.020	0.008	0.014	*	0.136
TTPA	0.001	0.008	0.934		0.005
<u>Post-hoc estimates by domain</u>					
Early numeracy					
TT	0.017	0.009	0.057	+	0.090
TTPA	-0.011	0.009	0.218		-0.062
Early literacy					
TT	0.015	0.012	0.218		0.073
TTPA	0.010	0.012	0.406		0.049
Social-emotional					
TT	0.032	0.012	0.008	**	0.166
TTPA	0.013	0.012	0.283		0.066
Executive function					
TT	0.016	0.012	0.168		0.088
TTPA	0.008	0.011	0.503		0.043
<u>Approaches to learning</u>					
TT	0.073	0.039	0.061	+	0.011
TTPA	0.020	0.029	0.612		0.030
Sample size = 2,975					

Notes. Estimates are computed using observed scores, in three level models: children nested in classrooms nested in schools. Effect sizes calculated accounting for the 3-level model structure (Hedges, 2009).

Sample includes children present at baseline and follow-up.

TT = Teacher training condition; TTPA = teacher training plus parent awareness training condition.

Models include the following control variables: private (vs. public) sector status of the school, six district dummies, a dummy variable for if the school was assigned to receive teacher text messages, a dummy for if the school was assigned to receive parent flyers, a series of five dummy variables accounting for within-sample mobility, child gender, age, KG level (1, 2, or 3 if KG1 and KG2 were combined in one classroom, as a categorical variable), and baseline score for each respective outcome.

Table 13. Moderation of Treatment Impacts on Children's School Readiness Composite and Individual Domains, by Grade Level

	School readiness composite	Early Numeracy	Early Literacy	Social-emotional	Executive Function
TT	0.024 * (.010)	0.013 (.011)	0.020 (.014)	0.035 * (.015)	0.023 + (.014)
TTPA	0.008 (.010)	-0.007 (.011)	-0.002 (.014)	0.021 (.015)	0.016 (.014)
Grade level (1=KG1, 2=KG2)	0.043 *** (.007)	0.039 *** (.009)	0.064 *** (.011)	0.052 *** (.013)	0.068 *** (.011)
Grade level*TT	-0.013 (.009)	0.003 (.012)	-0.019 (.014)	-0.011 (.017)	-0.022 (.015)
Grade level*TTPA	-0.019 * (.009)	-0.012 (.012)	-0.022 + (.013)	-0.021 (.017)	-0.020 (.015)

Note. + $p < .10$.

Estimates are computed using observed scores, in three level models: children nested in classrooms nested in schools. Effect sizes calculated accounting for the 3-level model structure (Hedges, 2009).

Sample includes children present at baseline and follow-up.

TT = Teacher training condition; TTPA = teacher training plus parent awareness training condition.

Models include the following control variables: private (vs. public) sector status of the school, six district dummies, a dummy variable for if the school was assigned to receive teacher text messages, a dummy for if the school was assigned to receive parent flyers, a series of five dummy variables accounting for within-sample mobility, child gender, age, KG level (1, 2, or 3 if KG1 and KG2 were combined in one classroom, as a categorical variable), and baseline score for each respective outcome.

Table 14. Moderation of Treatment Impacts on Outcomes, by Public and Private Sector Status

	b	se	p-value	
<u>Teacher professional well-being</u>				
Motivation				
TT	0.12	0.09	0.183	
TTPA	-0.01	0.09	0.875	
Private school	0.02	0.11	0.833	
TT*Private	-0.08	0.11	0.478	
TTPA*Private	-0.04	0.12	0.733	
Burnout				
TT	-0.08	0.16	0.619	
TTPA	-0.31	0.16	0.059	+
Private school	-0.01	0.19	0.976	
TT*Private	-0.43	0.21	0.042	*
TTPA*Private	-0.32	0.21	0.129	
Job Satisfaction				
TT	-0.23	0.13	0.078	+
TTPA	-0.13	0.13	0.328	
Private school	-0.29	0.15	0.050	*
TT*Private	0.29	0.17	0.086	+
TTPA*Private	0.27	0.17	0.107	
Teacher turnover				
TT	-0.40	0.67	0.549	
TTPA	0.07	0.64	0.909	
Private school	1.70	0.69	0.013	*
TT*Private	-1.32	0.77	0.089	+
TTPA*Private	-1.38	0.78	0.078	+
<u>Classroom processes</u>				
Facilitating Deeper Learning				
TT	0.03	0.14	0.812	
TTPA	-0.03	0.14	0.851	
Private school	0.16	0.16	0.309	
TT*Private	-0.04	0.18	0.830	
TTPA*Private	-0.06	0.18	0.744	
Emotional Support & Behavior Management				
TT	0.18	0.07	0.016	*
TTPA	0.18	0.07	0.017	*
Private school	0.16	0.09	0.061	+
TT*Private	0.04	0.10	0.690	

	b	se	p-value	
<u>Teacher professional well-being</u>				
TTPA*Private	0.00	0.10	0.966	
<u>Supporting Student Expression</u>				
TT	0.42	0.13	0.002	**
TTPA	0.01	0.13	0.962	
Private school	0.06	0.16	0.701	
TT*Private	-0.18	0.17	0.307	
TTPA*Private	0.17	0.18	0.335	
<u>Child school readiness composite</u>				
TT	0.011	0.011	0.312	
TTPA	-0.003	0.011	0.761	
Private school	0.022	0.010	0.031	*
TT*Private	0.016	0.014	0.231	
TTPA*Private	0.008	0.014	0.565	

Notes. Estimates for teacher / classroom level outcomes are computed using observed scores, in two level models: teachers nested in schools. Sample includes only teachers present at baseline and follow up. Estimates for child outcomes are computed using observed scores, in three-level models: children nested in classrooms nested in schools. Sample includes only teachers present at baseline and follow up.

Models include the following control variables: private (vs. public) sector status of the school, six district dummies, a dummy variable for if the school was assigned to receive teacher text messages, a dummy for if the school was assigned to receive parent flyers, a series of five dummy variables accounting for within-sample mobility. For child outcomes, we also included child gender, age, KG level (1, 2, or 3 if KG1 and KG2 were combined in one classroom, as a categorical variable), and baseline score for each respective outcome. For teacher outcomes, we also included teacher gender, age, level of education, years of teaching experience, and baseline score for each respective outcome.

Table 8a. Impact Coefficients for Teacher Professional Well-Being Outcomes with TT as the Reference Group

VARIABLES	Motivation	Burnout	Job satisfaction
1.tx (TTPA vs. TT)	-0.109 (0.076)	-0.166 (0.140)	-0.089 (0.112)
2.tx (Control vs. TT)	-0.078 (0.069)	0.286** (0.125)	-0.089 (0.100)
Observations	344	344	344
Number of groups	211	211	211

Standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Notes. Estimates are computed using observed scores, in two level models: teachers nested in schools. Effect sizes calculated accounting for the 2-level model structure (Hedges, 2009).

Sample includes teachers present at baseline and follow-up.

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

TT = Teacher training condition; TTPA = Teacher training plus parent awareness training condition.

Models include the following control variables: private (vs. public) sector status of the school, six district dummies, a dummy variable for if the school was assigned to receive teacher text messages, a dummy for if the school was assigned to receive parent flyers, a series of five dummy variables accounting for within-sample mobility, teacher gender, age, level of education, years of teaching experience, the baseline score for each respective outcome.

Table 8b. Impact Coefficients for Classroom Quality Outcomes with TT as the Reference Group

VARIABLES	Facilitating Deeper Learning	Emotional Support & Behavior Management	Supporting Student Expression
1.tx (TTPA vs. TT)	-0.0687 (0.117)	-0.0230 (0.0625)	-0.230** (0.115)
2.tx (Control vs. TT)	-0.0156 (0.107)	-0.196*** (0.0570)	-0.321*** (0.105)
Observations	337	337	337
Number of groups	205	205	205

Standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Notes. Estimates are computed using observed scores, in two level models: teachers nested in schools. Effect sizes calculated accounting for the 2-level model structure (Hedges, 2009).

Sample includes teachers present at baseline and follow-up.

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

TT = Teacher training condition; TTPA = Teacher training plus parent awareness training condition.

Models include the following control variables: private (vs. public) sector status of the school, six district dummies, a dummy variable for if the school was assigned to receive teacher text messages, a dummy for if the school was assigned to receive parent flyers, a series of five dummy variables accounting for within-sample mobility, teacher gender, age, level of education, and years of teaching experience.

Table 9. Impact Coefficients for Child Outcomes with TT as the Reference Group

VARIABLES	School readiness composite	Early numeracy	Early literacy	Social- emotional	Executive function	Approaches to Learning
1.tx (TTPA vs. TT)	-0.020** (0.009)	-0.028*** (0.010)	-0.024* (0.013)	-0.019 (0.014)	-0.008 (0.013)	0.0727* (0.0387)
2.tx (Control vs. TT)	-0.020** (0.008)	-0.017* (0.009)	-0.015 (0.012)	-0.032*** (0.012)	-0.016 (0.012)	0.0196 (0.0387)
Observations	2,975	2,975	2,975	2,975	2,975	2,975
Number of groups	235	235	235	235	235	235

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Estimates are computed using observed scores, in three level models: children nested in classrooms nested in schools. Effect sizes calculated accounting for the 3-level model structure (Hedges, 2009).

Sample includes children present at baseline and follow-up.

TT = Teacher training condition; TTPA = teacher training plus parent awareness training condition.

Models include the following control variables: private (vs. public) sector status of the school, six district dummies, a dummy variable for if the school was assigned to receive teacher text messages, a dummy for if the school was assigned to receive parent flyers, a series of five dummy variables accounting for within-sample mobility, child gender, age, KG level (1, 2, or 3 if KG1 and KG2 were combined in one classroom, as a categorical variable), and baseline score for each respective outcome.

Table 10. Impact Estimates for Professional Well-Being and Classroom Quality Outcomes Using Only the Stratification Variables as Covariates

	Teacher professional well-being					Classroom quality						
	Motivation		Burnout		Job satisfaction	FDL		ESBM		SSE		
1.tx (TT vs. control)	0.110	+	-0.314	**	0.080	0.046		0.205	***	0.231	**	
	(.059)		(.120)		(.097)	(.092)		(.048)		(.090)		
2.tx (TTPA vs. control)	-0.004		-0.193		0.018	-0.052		0.169	**	0.125		
	(.060)		(.121)		(.098)	(.093)		(.049)		(.091)		
private	0.082	+	-0.226	*	0.019	-0.015		0.055		0.046		
	(.049)		(.099)		(.080)	(.076)		(.039)		(.074)		
2.district	-0.129	+	-0.325	*	0.096	0.439	***	0.228	***	0.408	***	
	(.075)		(.152)		(.124)	(.144)		(.061)		(0.115)		
3.district	0.074		-0.318	+	-0.274	+	0.242	+	0.122	+	0.211	+
	(.085)		(.172)		(.140)		(.130)		(.068)		(.127)	
4.district	-0.012		-0.143		-0.132	0.416	***	-0.029		0.446	***	
	(.076)		(.154)		(.125)	(.119)		(.062)		(.116)		
5.district	-0.145	+	0.0652		-0.169	0.254	+	0.189	**	0.075		
	(.087)		(.177)		(.144)	(.136)		(.071)		(.133)		
6.district	-0.166	+	-0.304		0.169	0.434	**	0.196	**	0.456	***	
	(.094)		(.191)		(.155)	(.145)		(.076)		(.142)		
Constant	4.689	***	2.484	***	3.089	***	2.085	***	2.795	***	1.315	***
	(.071)		(.144)		(.117)	(.110)		(.058)		(.108)		
Observations	347		347		347	340		340		340		
Number of groups	212		212		212	206		206		206		

Standard errors in parentheses

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, + $p < .10$.

Notes. Estimates are computed using observed scores, in two level models: teachers nested in schools. Effect sizes calculated accounting for the 2-level model structure (Hedges, 2009).

Sample includes teachers present at baseline and follow-up.

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

TT = Teacher training condition; TTPA = Teacher training plus parent awareness training condition.

Models include the following control variables: private (vs. public) sector status of the school, six district dummies, a dummy variable for if the school was assigned to receive teacher text messages, a dummy for if the school was assigned to receive parent flyers, a series of five dummy variables accounting for within-sample mobility, teacher gender, age, level of education, and years of teaching experience. Models for teacher professional well-being control for baseline values for each respective outcome.

Table 11. Impact Estimates for Child Outcomes Using Only the Stratification Variables as Covariates

	School readiness		Numeracy		Literacy		Social- emotional		Exec. Function		Approaches to learning	
1.tx (TT vs. control)	0.021	+	0.019		0.017		0.033	**	0.016		0.080	*
	(.012)		(.014)		(.015)		(.012)		(.012)		(.037)	
2.tx (TTPA vs. control)	0.0146		0.011		0.007		0.029	*	0.011		0.065	+
	(.012)		(.014)		(.015)		(.012)		(.011)		(.037)	
private	0.036	***	0.033	**	0.095	***	-0.014		0.032	***	0.080	***
	(.010)		(.012)		(.013)		(.010)		(.010)		(.030)	
2.district	0.041	**	0.028		0.038	*	0.077	***	0.021		-0.700	***
	(.015)		(.019)		(.020)		(.015)		(.015)		(.048)	
3.district	0.079	***	0.065	**	0.101	***	0.089	***	0.063	***	-0.344	***
	(.018)		(.022)		(.022)		(.018)		(.017)		(.055)	
4.district	-0.005		0.016		-0.002		0.021		-0.056	***	-0.567	***
	(.016)		(.019)		(.020)		(.016)		(.015)		(.049)	
5.district	0.067	***	0.069	**	0.107	***	0.065	***	0.029	+	0.0136	
	(.018)		(.022)		(.023)		(.018)		(.018)		(.056)	
6.district	-0.002		0.003		0.005		0.014		-0.029		-0.631	***
	(.018)		(.023)		(.024)		(.019)		(.018)		(.058)	
Constant	0.506	***	0.524	***	0.524	***	0.413	***	0.562	***	3.541	***
	(.015)		(.018)		(.019)		(.015)		(.014)		(.046)	
Observations	2,975		2,975		2,975		2,975		2,975		2,975	
Number of groups	235		235		235		235		235		235	

Standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05, + p<.10.

Notes. Estimates are computed using observed scores, in three level models: children nested in classrooms nested in schools. Effect sizes calculated accounting

for the 3-level model structure (Hedges, 2009).

Sample includes children present at baseline and follow-up.

TT = Teacher training condition; TTPA = teacher training plus parent awareness training condition.

Models include the following control variables: private (vs. public) sector status of the school, six district dummies, a dummy variable for if the school was assigned to receive teacher text messages, a dummy for if the school was assigned to receive parent flyers, a series of five dummy variables accounting for within-sample mobility, child gender, age, KG level (1, 2, or 3 if KG1 and KG2 were combined in one classroom, as a categorical variable), and baseline score for each respective outcome.

Table 12. Attrition from Baseline to Midline for Children and Teachers, by Treatment Condition

	Children		Teachers	
	Baseline	Midline	Baseline	Midline
Sample size				
Control	1,180	931	139	97
TT	1,167	1,025	155	128
TTPA	1,088	1,019	150	122
Total	3,435	2,975	444	347

Table 13. Upper and Lower Bound Estimates for Impact Estimates on Teacher Professional Well-Being.

	Original estimates				Lower bound estimates (v1)				Lower bound estimates (v2)				Upper bound estimates			
	b	se	p-value	<i>e.s.</i> (<i>d_{av}</i>)	b	se	p-value	<i>e.s.</i> (<i>d_{av}</i>)	b	se	p-value	<i>e.s.</i> (<i>d_{av}</i>)	b	se	p-value	<i>e.s.</i> (<i>d_{av}</i>)
<i>Motivation</i>																
Teacher	0.078	0.068	0.256	0.180	0.096	0.064	0.131	0.219	0.057	0.061	0.349	0.139	0.121	0.063	0.054 +	0.14
Teacher + Parent	-0.031	0.071	0.660	-0.071	0.006	0.064	0.926	0.014	0.028	0.063	0.653	0.068	0.048	0.064	0.447	0.055
<i>Burnout</i>																
Teacher	-0.286	0.125	0.022 *	0.321	-0.270	0.126	0.032 *	0.304	-0.237	0.113	0.036 *	0.284	-0.412	0.118	0.000 ***	0.476
Teacher + Parent	-0.452	0.130	0.000 **	0.507	-0.315	0.120	0.014 *	0.354	-0.368	0.116	0.001 ***	0.441	-0.551	0.120	0.000 ***	0.637
<i>Job Satisfaction</i>																
Teacher	-0.089	0.100	0.375	-0.131	0.143	0.105	0.174	0.210	0.012	0.094	0.898	0.018	0.174	0.098	0.076 +	0.262
Teacher + Parent	0.000	0.100	0.999	0.000	0.175	0.107	0.103	0.258	-0.085	0.095	0.375	-0.129	0.133	0.010	0.183	0.201
	Sample size = 344				Sample size = 397				Sample size = 398				Sample size = 398			

Notes. Estimates are computed using observed scores, in two level models: teachers nested in schools. Effect sizes calculated accounting for the 2-level model structure (see Hedges). All models control for covariates specified in the report.

Lower bound v1 estimates calculated by using propensity score matching based on baseline characteristics for treatment teachers who attrited, and imputing their follow up score based on their matched control group pair; lower bound v2 estimates calculated by imputing the follow up score the treatment teachers who attrited as the 25th percentile score of treatment teachers who were present at follow up in their respective treatment condition.

Upper bound estimates calculated by imputing the follow up score for treatment teachers who attrited as the 75th percentile score of treatment teachers who were present at follow up in their respective treatment condition.

Table 14. Upper and Lower Bound Estimates for Impact Estimates on Classroom Quality.

	Original estimates				Lower bound estimates (v1)				Lower bound estimates (v2)				Upper bound estimates			
	b	se	p-value	e.s. (d_{wt})	b	se	p-value	e.s. (d_{wt})	b	se	p-value	e.s. (d_{wt})	b	se	p-value	e.s. (d_{wt})
<i>Facilitating Deeper Learning</i>																
Teacher	0.016	0.107	0.880	0.025	-0.005	0.105	0.959	-0.008	-0.024	0.100	0.809	-0.038	0.094	0.101	0.356	0.149
Teacher + Parent	-0.052	0.109	0.663	-0.081	-0.138	0.107	0.196	-0.210	-0.114	0.101	0.261	-0.182	-0.018	0.102	0.859	-0.029
<i>Emotional Support & Behavior Management</i>																
Teacher	0.196	0.057	0.001 ***	0.520	0.172	0.057	0.002 **	0.458	0.181	0.053	0.001 ***	0.510	0.229	0.053	0.000	0.621
Teacher + Parent	0.173	0.059	0.003 **	0.459	0.103	0.057	0.073 +	0.274	0.123	0.054	0.023 *	0.346	0.205	0.054	0.000	0.556
<i>Supporting Student Expression</i>																
Teacher	0.321	0.106	0.002 **	0.499	0.272	0.010	0.007 **	0.413	0.245	0.098	0.012 *	0.378	0.388	0.101	0.000	0.589
Teacher + Parent	0.092	0.109	0.398	0.143	0.082	0.101	0.416	0.125	0.015	0.010	0.881	0.023	0.160	0.102	0.117	0.243
	Sample size = 337				Sample size = 379				Sample size = 380				Sample size = 380			

Notes. Estimates are computed using observed scores, in two level models: teachers nested in schools. Effect sizes calculated accounting for the 2-level model structure (see Hedges).

Lower bound v1 estimates calculated by using propensity score matching based on baseline characteristics for treatment teachers who attrited, and imputing their follow up score based on their matched control group pair; lower bound v2 estimates calculated by imputing the follow up score the treatment teachers who attrited as the 25th percentile score of treatment teachers who were present at follow up in their respective treatment condition.

Upper bound estimates calculated by imputing the follow up score for treatment teachers who attrited as the 75th percentile score of treatment teachers who were present at follow up in their respective treatment condition.

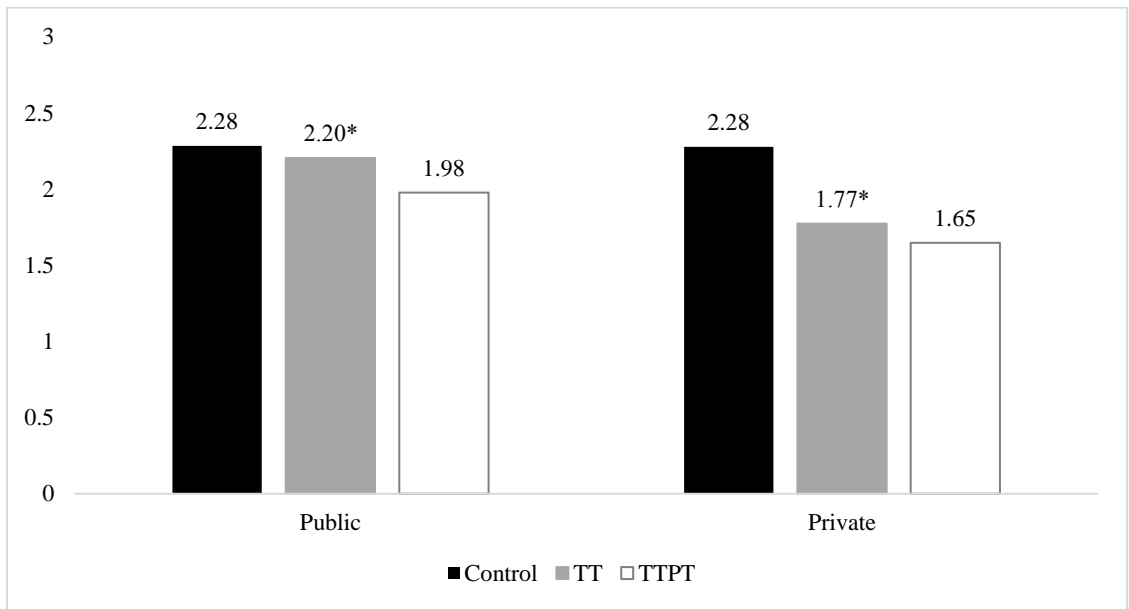


Figure 2. Moderation on Teacher Burnout by Public and Private Schools

* indicates that the interaction between TT treatment status, private sector status, and teacher burnout is statistically significant at $p < .05$.

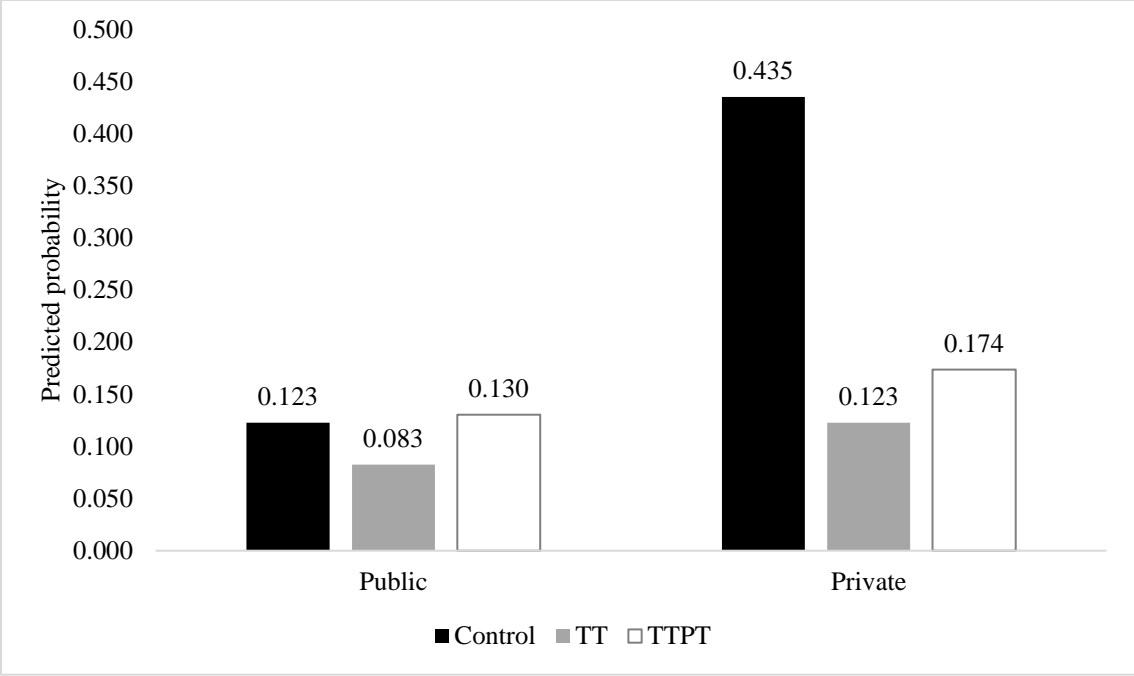


Figure 3. Predicted Probability of Teacher Turnover by Treatment Status for Public and Private Schools

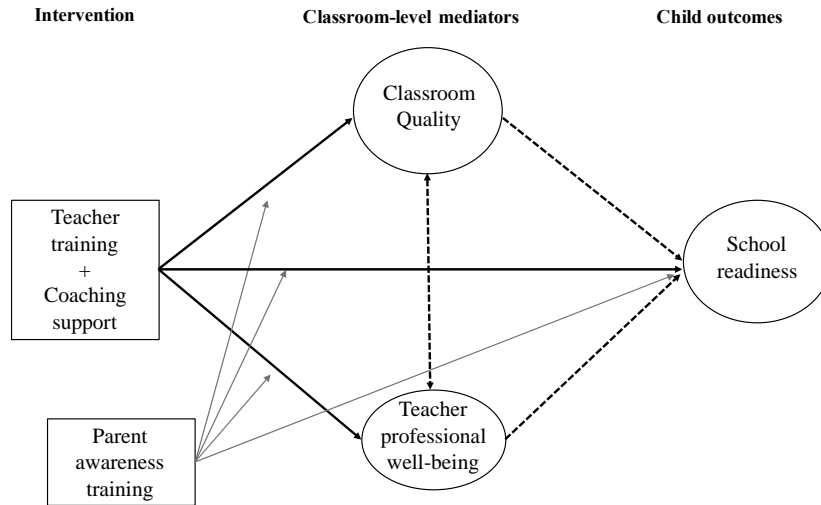


Figure 4. QP4G Theory of Change

Notes. Solid lines represent causal relationships. Dashed lines represent non-causal relationship

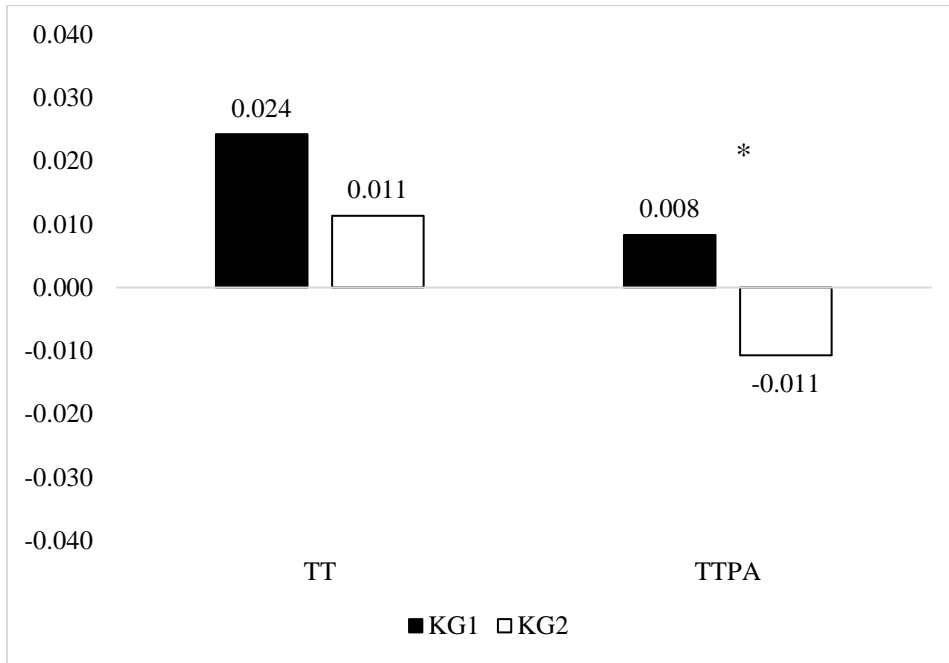


Figure 5. Impact Estimates for TT and TTPA Treatment Conditions, by Child Grade Level

* indicates that the interaction between TTPA treatment status, child grade level, and children's school readiness composite score is statistically significant at $p < .05$.

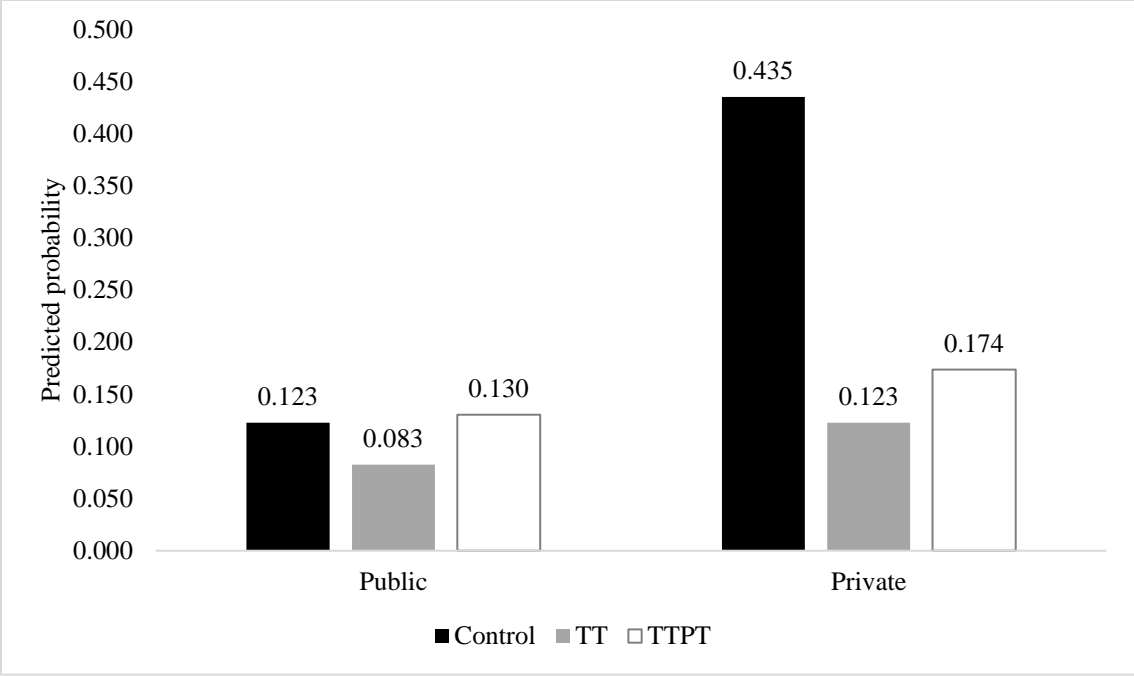


Figure 6. Predicted Probability of Teacher Turnover by Treatment Status for Public and Private Schools

Appendix A. Sensitivity of Impacts Analysis

While the baseline equivalency checks conducted across school, teacher, and child characteristics led to the conclusion that random assignment was successful, there were four baseline variables that were not equivalent across the three conditions. These were head teacher years of experience ($F = 3.24$, $p = .041$); head teacher wants to leave the education profession ($F = 2.96$, $p = .054$); and teachers' mother's education level ($F = 4.81$, $p = .009$) and father's education level ($F = 3.30$, $p = .038$).

It is plausible that these characteristics lead schools and/or teachers to respond to the intervention differently than others. As a sensitivity check, we re-ran all impact analyses including (a) these four additional covariates in all models, and (b) interaction term between these four additional covariates with treatment status. This allowed us to assess if these observed characteristics, which were not balanced across treatment groups, and their interaction with treatment, were obscuring the results of the study.

The first column in Appendix Table A1 shows the original results for child outcomes; the second column shows the results with the addition of these four covariates; and the third column shows the results with the addition of these four covariates interacted with treatment status. Adding in the controls does not change the impact estimates (column 2). When adding in the controls interacted with treatment status does, the direction of all of the coefficients is the same, and some coefficients increase in magnitude.

Appendix Table A2 shows the same for classroom quality outcomes. Adding in the controls does not change the impact estimates. When adding in the controls interacted with treatment status does, however, none of the impact estimates are statistically significant. It appears that the inclusion of these terms adds additional measurement error to the analysis. In some cases, results are no longer statistically significant because the standard errors become much larger, and in some cases the

coefficients are close to zero.

Overall we conclude that our original impact estimates were sound.

Appendix Table A1. Results of Sensitivity Analysis for Impacts on Child Outcomes

	Original	Add'l controls	Add'l controls interacted with treatment
	b (SE)		
School readiness composite			
TT	0.020*	0.020*	0.034+
	(.009)	(.008)	(.020)
TTPA	0.001	0.002	0.005
	(.008)	(.009)	(.021)
Early numeracy			
TT	0.017+	0.018*	-0.001
	(.009)	(.009)	(.021)
TTPA	-0.011	-0.012	-0.049*
	(.009)	(.009)	(.022)
Early literacy			
TT	0.015	0.015	0.018
	(.012)	(.012)	(.028)
TTPA	0.010	-0.007	-0.010
	(.012)	(.012)	(.029)
Social-emotional			
TT	0.032*	0.030*	0.065***
	(.012)	(.012)	(.028)
TTPA	0.013	0.016	0.011
	(.012)	(.012)	(.030)
Executive function			
TT	0.016	0.014	0.045+
	(.012)	(.011)	(.027)
TTPA	0.008	0.004	0.038
	(.011)	(.012)	(.028)
Sample size = 2,975 children			

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

Estimates are computed using observed scores, in three level models: children nested in classrooms nested in schools. Effect sizes calculated accounting for the 3-level model structure (Hedges, 2009).

Sample includes children present at baseline and follow-up.

TT = Teacher training condition; TTPA = teacher training plus parent awareness training condition. Models include the following control variables: private (vs. public) sector status of the school, six district dummies, a dummy variable for if the school was assigned to receive teacher text messages, a dummy for if the school was assigned to receive parent flyers, a series of five dummy variables accounting for within-sample mobility, child gender, age, KG level (1, 2, or 3 if KG1 and KG2 were combined in one classroom, as a categorical variable), and baseline score for each respective outcome.

Appendix Table A2. Results of Sensitivity Analysis for Impacts on Classroom Quality

	Original	Add'l controls	Add'l controls interacted with treatment
	b (SE)		
Facilitating Deeper Learning			
TT	0.016 (.011)	0.031 (.109)	-0.302 (.254)
TTPA	-0.052 (.110)	-0.077 (.114)	-0.409 (.266)
Emotional Support & Behavior Management			
TT	0.196*** (.057)	0.182*** (.058)	0.211 (.133)
TTPA	0.173** (.059)	0.171** (.061)	-0.002 (.141)
Supporting Student Expression			
TT	0.321** (.105)	0.325** (.107)	-0.066 (.249)
TTPA	0.092 (.108)	0.102 (.112)	-0.110 (.263)
Sample size = 337 classrooms			

Notes. Estimates are computed using observed scores, in two level models: teachers nested in schools. Effect sizes calculated accounting for the 2-level model structure (Hedges, 2009).

Sample includes teachers present at baseline and follow-up.

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

TT = Teacher training condition; TTPA = Teacher training plus parent awareness training condition.

Models include the following control variables: private (vs. public) sector status of the school, six district dummies, a dummy variable for if the school was assigned to receive teacher text messages, a dummy for if the school was assigned to receive parent flyers, a series of five dummy variables accounting for within-sample mobility, teacher gender, age, level of education, and years of teaching experience.

Appendix B. Impact Estimate Models Output with Covariates

This section includes the main impact models shown above, with the full model output including coefficients for all covariates.

Table B1. Impact Estimates with Covariates for Teacher Professional Well-Being Outcomes

VARIABLES	Motivation	Burnout	ob satisfaction
1.tx (TT)	0.0778 (0.0685)	-0.286** (0.125)	-0.0888 (0.1000)
2.tx (TTPA)	-0.0312 (0.0710)	-0.452*** (0.130)	0.000136 (0.103)
teacher_text	0.0475 (0.0760)	0.00215 (0.139)	0.134 (0.113)
parent_text	-0.0230 (0.110)	0.388* (0.200)	-0.146 (0.161)
Baseline score	0.143*** (0.0397)	0.430*** (0.0457)	0.456*** (0.0536)
Teacher age	3.48e-05 (0.00282)	-0.00778 (0.00515)	-0.00646 (0.00395)
2.edlevel	-0.0124 (0.111)	-0.134 (0.201)	0.147 (0.155)
3.edlevel	-0.0452 (0.114)	-0.278 (0.208)	0.0830 (0.160)
4.edlevel	-0.164 (0.125)	-0.226 (0.229)	0.0998 (0.176)
Female	0.0813 (0.161)	0.411 (0.294)	0.379* (0.225)
Years teaching	-0.00170 (0.00422)	0.00407 (0.00772)	-0.000254 (0.00591)
Private school	-0.0223 (0.0749)	-0.278** (0.136)	-0.0907 (0.106)
2.district	-0.121 (0.0773)	-0.161 (0.141)	-0.0108 (0.113)
3.district	0.0687 (0.0849)	-0.238 (0.154)	0.316** (0.124)
4.district	-0.00443 (0.0767)	-0.000779 (0.140)	0.167 (0.112)
5.district	-0.164* (0.0891)	0.206 (0.162)	0.220* (0.129)
6.district	-0.215** (0.0967)	-0.138 (0.174)	-0.0214 (0.139)
dummy_sch_merge	-0.183 (0.195)	0.804** (0.356)	0.531* (0.280)
dummy_sch_split	0.234 (0.258)	0.372 (0.469)	-0.0407 (0.374)
dummy_comb_spl	-0.00440 (0.174)	0.119 (0.318)	0.686*** (0.246)
dummy_spl_comb	0.243 (0.210)	0.213 (0.385)	0.186 (0.298)
dummy_contam	0.139 (0.309)	0.111 (0.565)	0.513 (0.457)
Constant	4.058*** (0.290)	1.577*** (0.447)	0.771** (0.348)
Observations	344	344	344
Number of groups	211	211	211

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes. Estimates are computed using observed scores, in two level models: teachers nested in schools. Effect sizes calculated accounting for the 2-level model structure (Hedges, 2009).

Sample includes teachers present at baseline and follow-up.

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

TT = Teacher training condition; TTPA = Teacher training plus parent awareness training condition.

Models include the following control variables: private (vs. public) sector status of the school, six district dummies, a dummy variable for if the school was assigned to receive teacher text messages, a dummy for if the school was assigned to receive parent flyers, a series of five dummy variables accounting for within-sample mobility, teacher gender, age, level of education, years of teaching experience, and the baseline value for each respective outcome.

Table B2. Impact Estimates with Covariates for Classroom Quality Outcomes

VARIABLES	Emotional Support		
	Facilitating Deeper Learning	& Behavioral Management	Supporting Student Expression
1.tx (TT)	0.0156 (0.107)	0.196*** (0.0570)	0.321*** (0.105)
2.tx (TTPA)	-0.0531 (0.110)	0.173*** (0.0587)	0.0915 (0.108)
teacher_text	-0.0497 (0.117)	-0.0229 (0.0623)	-0.260** (0.115)
parent_text	-0.00210 (0.168)	0.0284 (0.0897)	0.266 (0.166)
Teacher age	-0.00463 (0.00413)	-0.00468** (0.00230)	0.00250 (0.00426)
2.edlevel	-0.190 (0.161)	0.0416 (0.0900)	-0.131 (0.167)
3.edlevel	0.00582 (0.166)	0.202** (0.0927)	-0.113 (0.172)
4.edlevel	0.130 (0.183)	0.312*** (0.102)	-0.119 (0.190)
Female	-0.228 (0.234)	0.0954 (0.131)	-0.189 (0.242)
Years teaching	-0.0115* (0.00617)	0.000760 (0.00344)	-0.0140** (0.00637)
Private school	0.129 (0.111)	0.177*** (0.0608)	0.0542 (0.112)
2.district	0.468*** (0.118)	0.220*** (0.0631)	0.440*** (0.117)
3.district	0.293** (0.129)	0.110 (0.0688)	0.253** (0.127)
4.district	0.492*** (0.118)	-0.0253 (0.0631)	0.485*** (0.117)
5.district	0.317** (0.137)	0.191*** (0.0735)	0.0703 (0.136)
6.district	0.533*** (0.146)	0.219*** (0.0780)	0.501*** (0.144)
dummy_sch_merge	0.131 (0.293)	0.0673 (0.159)	-0.494* (0.294)
dummy_sch_split	0.994 (0.624)	0.0655 (0.345)	0.391 (0.638)
dummy_comb_spl	0.329 (0.256)	0.00848 (0.142)	0.210 (0.262)
dummy_spl_comb	-0.191 (0.312)	-0.0426 (0.172)	-0.372 (0.318)
dummy_contam	0.234 (0.479)	0.0476 (0.253)	-0.109 (0.467)
Constant	2.422*** (0.346)	2.663*** (0.192)	1.527*** (0.356)
Observations	337	337	337
Number of groups	205	205	205

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes. Estimates are computed using observed scores, in two level models: teachers nested in schools. Effect sizes

calculated accounting for the 2-level model structure (Hedges, 2009).

Sample includes teachers present at baseline and follow-up.

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

TT = Teacher training condition; TTPA = Teacher training plus parent awareness training condition.

Models include the following control variables: private (vs. public) sector status of the school, six district dummies, a dummy variable for if the school was assigned to receive teacher text messages, a dummy for if the school was assigned to receive parent flyers, a series of five dummy variables accounting for within-sample mobility, teacher gender, age, level of education, and years of teaching experience.

Table B3. Impact Estimates with Covariates for Child School Readiness Outcomes

VARIABLES	School readiness	Early numeracy	Early literacy	Social- emotional	Executive function
1.tx (TT)	0.0205** (0.00836)	0.0167* (0.00878)	0.0145 (0.0118)	0.0322*** (0.0121)	0.0159 (0.0115)
2.tx (TTPA)	0.000916 (0.00835)	-0.0108 (0.00876)	-0.00976 (0.0118)	0.0130 (0.0121)	0.00769 (0.0115)
Baseline score	0.529*** (0.0129)	0.595*** (0.0143)	0.517*** (0.0150)	0.317*** (0.0173)	0.238*** (0.0118)
parent_text	0.0143 (0.0137)	0.0260* (0.0143)	0.00986 (0.0193)	0.0197 (0.0198)	0.00328 (0.0188)
teacher_text	-0.00556 (0.00955)	-0.0128 (0.0100)	-0.00538 (0.0135)	-0.00253 (0.0138)	0.00132 (0.0132)
2.KG	0.0315*** (0.00460)	0.0353*** (0.00581)	0.0497*** (0.00665)	0.0398*** (0.00762)	0.0519*** (0.00693)
3.KG	0.00232 (0.0126)	0.00679 (0.0140)	0.00197 (0.0180)	0.00411 (0.0194)	0.0172 (0.0183)
private	0.0301*** (0.00611)	0.0246*** (0.00647)	0.0596*** (0.00869)	0.00183 (0.00887)	0.0395*** (0.00842)
chmale	-0.00527 (0.00353)	0.00240 (0.00445)	-0.0104** (0.00480)	-0.0137** (0.00634)	-0.000614 (0.00562)
childage_I	0.00667*** (0.00176)	0.00945*** (0.00221)	0.00685*** (0.00238)	0.0120*** (0.00307)	0.0116*** (0.00274)
2.district	0.0539*** (0.00940)	0.0371*** (0.00986)	0.0464*** (0.0132)	0.0822*** (0.0136)	0.0335*** (0.0129)
3.district	0.0628*** (0.0105)	0.0452*** (0.0110)	0.0840*** (0.0148)	0.0739*** (0.0152)	0.0610*** (0.0144)
4.district	-0.00336 (0.00937)	0.0223** (0.00983)	0.00761 (0.0132)	0.0201 (0.0136)	-0.0568*** (0.0129)
5.district	0.0334*** (0.0108)	0.0346*** (0.0114)	0.0646*** (0.0153)	0.0509*** (0.0157)	0.0133 (0.0149)
6.district	-0.00543 (0.0111)	0.000787 (0.0116)	2.34e-05 (0.0157)	0.00389 (0.0161)	-0.0205 (0.0153)
dummy_sch_merge	-0.0394** (0.0201)	-0.0157 (0.0210)	-0.0484* (0.0283)	-0.0142 (0.0290)	-0.0372 (0.0276)
dummy_sch_split	0.0754** (0.0321)	0.0725** (0.0339)	0.104** (0.0453)	0.00348 (0.0468)	0.107** (0.0444)
dummy_comb_spl	0.0233 (0.0154)	0.00720 (0.0164)	0.0177 (0.0219)	0.0317 (0.0226)	0.0325 (0.0214)
dummy_spl_comb	0.0128 (0.0205)	0.00721 (0.0219)	0.0476* (0.0287)	-0.00339 (0.0303)	0.0104 (0.0286)
dummy_contam	-0.0172 (0.0427)	-0.0530 (0.0448)	0.0355 (0.0602)	0.0463 (0.0618)	-0.0509 (0.0587)
Constant	0.215*** (0.0135)	0.196*** (0.0155)	0.261*** (0.0185)	0.195*** (0.0219)	0.330*** (0.0201)
Observations	2,975	2,975	2,975	2,975	2,975
Number of groups	235	235	235	235	235

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes. Estimates are computed using observed scores, in three level models: children nested in classrooms nested in schools. Effect sizes calculated accounting for the 3-level model structure (Hedges, 2009).

Sample includes children present at baseline and follow-up.

TT = Teacher training condition; TTPA = teacher training plus parent awareness training condition.

Models include the following control variables: private (vs. public) sector status of the school, six district dummies, a dummy variable for if the school was assigned to receive teacher text messages, a dummy for if the school was assigned to receive parent flyers, a series of five dummy variables accounting for within-sample mobility, child gender, age, KG level (1, 2, or 3 if KG1 and KG2 were combined in one classroom, as a categorical variable), and baseline score for each respective outcome.