University of Pennsylvania
ScholarlyCommons

Publicly Accessible Penn Dissertations

2020

Examining Teacher Performance In Ghana

Syeda Farwa Fatima
University of Pennsylvania

Follow this and additional works at: https://repository.upenn.edu/edissertations

 Part of the Statistics and Probability Commons, and the Teacher Education and Professional Development Commons

Recommended Citation

Fatima, Syeda Farwa, "Examining Teacher Performance In Ghana" (2020). *Publicly Accessible Penn Dissertations*. 3808.
https://repository.upenn.edu/edissertations/3808

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/edissertations/3808
For more information, please contact repository@pobox.upenn.edu.

Examining Teacher Performance In Ghana

Abstract

This study used longitudinal data on 444 teachers and 3,435 students to examine teacher performance in Ghana. The study is divided into two parts. The first part of the study examined factors that mediate the causal effects of a kindergarten teacher training program on classroom quality and student outcomes. Specifically, it examined whether teachers' knowledge of the learning content, teachers' implementation quality of behavioral and instructional practices and teachers' professional well-being were significant mediators of the treatment effect. It utilized a causal mediation approach, which allowed the average causal mediation effects to be parametrically and nonparametrically identified under a set of minimum conditions. The study found that implementation quality was a significant mediator of positive treatment effect on classroom quality across time. This effect persisted even when teacher knowledge and professional well-being were accounted for. The study also found small marginal mediation effects on student outcomes, including a positive mediation effect on literacy and a negative mediation effect on executive functioning in the presence of all mediators. Overall, this study provides empirical evidence to design future interventions that place more emphasis on the influential pathway of implementation quality to yield positive impacts, particularly in early education contexts.

The second part of the study examined teacher profiles that provide diagnostic information about teachers' instructional strengths and weaknesses. It applied stage-wise cluster analysis to reveal different subpopulations of teachers and study how they relate to student outcomes. The study found six profiles of teachers with varying professional well-being and classroom practices, including two that were significantly associated with positive student learning across all four domains of numeracy, literacy, socioemotional development and executive functioning. Overall, the results allow easy identification of growth opportunities for each profile of teachers that helps provide formative feedback and targeted support to facilitate high quality teaching and maximize positive student learning outcomes.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Education

First Advisor

Robert F. Boruch

Keywords

causal mediation, cluster analysis, Ghana, sub-Saharan Africa, teacher profiles, teacher training

Subject Categories

Education | Statistics and Probability | Teacher Education and Professional Development

This dissertation is available at ScholarlyCommons: https://repository.upenn.edu/edissertations/3808

EXAMINING TEACHER PERFORMANCE IN GHANA

Syeda Farwa Fatima

A DISSERTATION

in

Education

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2020

Supervisor of Dissertation

_____

Robert Boruch

University Trustee Chair Professor of Education and Statistics

Graduate Group Chairperson

_____

J. Matthew Hartley, Professor of Education

Dissertation Committee

Sharon Wolf, Assistant Professor of Education

Paul McDermott, Professor of Education

Michael Rovine, Senior Fellow

EXAMINING TEACHER PERFORMANCE IN GHANA

COPYRIGHT

2020

Syeda Farwa Fatima

This work is licensed under the
Creative Commons Attribution-
NonCommercial-NoDerivatives 4.0
International License

To view a copy of this license, visit the following pages:
https://creativecommons.org/licenses/by-nc-nd/4.0/
https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode

DEDICATION

To my loving brother, Zain.
For inspiring me to be a better person every day.

ACKNOWLEDGEMENTS

There are so many people I need to thank without whom I would not have been able to complete this dissertation. First, I would like to extend special thanks to my committee members for their invaluable support throughout the process. I would like to especially thank my advisor, Dr. Robert Boruch, for pushing me towards new challenging opportunities in research and teaching throughout these three years, for his humor and wit through stressful times, and for working through never-ending paperwork with me. I would like to thank Dr. Paul McDermott for all the confidence I gained learning so much from his classes, for being able to apply those learnings to practical research so easily, and for the nurturing and special attention he gave to help me succeed. I would like to thank Dr. Mike Rovine for patiently answering all my questions in every class, for always being ready to explain difficult concepts and share resources for methods that were completely foreign to me, yet I needed to learn. Finally, I would like to extend a very special thanks to Dr. Sharon Wolf, without whom I would not have been able to conduct this research – from helping me conceptualize the studies, better understand the results and publish this work for a wider audience.

I am also very grateful to other faculty members in the Statistics department at Wharton, especially those who provided excellent training through various classes and research projects. I am also grateful to my program manager, Christine Lee, for being a pillar of strength and helping me through all sorts of administrative work, and my program cohort for collaborating and learning together, for the helpful feedback on my work and for always being there for moral support.

Last but not the least, I am most grateful to my family, without whom I would not be here today. My parents have worked endlessly to provide me with the best education, which has allowed me to dream big and work hard. They also spent many months with me, taking over the housework, while I finished my PhD coursework and took my comprehensive exams. My constant rock and my husband, Hamza, has been instrumental to my success. He has helped me grow academically and professionally, all while providing endless emotional support and love. I am forever indebted to him to help me start and finish this journey. My older brother with special needs, Zain, who inspires me every day to work hard to make this world a better place. My older sister, brother-in-law and nephew – Sana, Shabeeh and Sajjad, who have been a huge support through my graduate education in the United States. My little brother, Aun, who has supported me with humor and love all the way from Pakistan. My loving grandmother, who has asked me about my dissertation every single time I've spoken to her in the past year and prayed for its completion. I am also grateful to numerous other family members, including my parents-in-law, cousins and friends who have supported and encouraged me in this journey. Thank you so much everyone – it really takes a village to be here!

ABSTRACT

EXAMINING TEACHER PERFORMANCE IN GHANA

Syeda Farwa Fatima

Robert Boruch

This study used longitudinal data on 444 teachers and 3,435 students to examine teacher performance in Ghana. The study is divided into two parts. The first part of the study examined factors that mediate the causal effects of a kindergarten teacher training program on classroom quality and student outcomes. Specifically, it examined whether teachers' knowledge of the learning content, teachers' implementation quality of behavioral and instructional practices and teachers' professional well-being were significant mediators of the treatment effect. It utilized a causal mediation approach, which allowed the average causal mediation effects to be parametrically and nonparametrically identified under a set of minimum conditions. The study found that implementation quality was a significant mediator of positive treatment effect on classroom quality across time. This effect persisted even when teacher knowledge and professional well-being were accounted for. The study also found small marginal mediation effects on student outcomes, including a positive mediation effect on literacy and a negative mediation effect on executive functioning in the presence of all mediators. Overall, this study provides empirical evidence to design future interventions that place more emphasis on the influential pathway of implementation quality to yield positive impacts, particularly in early education contexts.

The second part of the study examined teacher profiles that provide diagnostic information about teachers' instructional strengths and weaknesses. It applied stage-wise cluster analysis to reveal different subpopulations of teachers and study how they relate to student outcomes. The study found six profiles of teachers with varying professional well-being and classroom practices, including two that were significantly associated with positive student learning across all four domains of numeracy, literacy, socioemotional development and executive functioning. Overall, the results allow easy identification of growth opportunities for each profile of teachers that helps provide formative feedback and targeted support to facilitate high quality teaching and maximize positive student learning outcomes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

EXECUTIVE SUMMARY

There is increasing evidence of a "learning crisis" in the field of international education. While countries have taken concrete steps to increase access to education, the quality of education provided is substandard in most developing countries. Data from PISA for Development (PISA-D), a new initiative to include low- and middle-income countries in the assessment, revealed disturbing results. Among the seven countries that participated, only 12% of the children were proficient in math and 23% in reading, compared to 77% and 80% in Organization for Economic Cooperation and Development (OECD) countries, respectively (Kaffenberger, 2019). The World Development Report 2018 identifies three main dimensions of the learning crisis: poor learning outcomes of children, lack of trained and motivated teachers and classroom facilities, and deeper systemic causes characterized by low accountability and high inequality (World Bank, 2018). Other research shows that the learning crisis is in fact, a reflection of the "teaching crisis" (Bold et al., 2017). Studies in developing countries illustrate the need for strong teachers to help students learn better (Buhl-Wiggers et al., 2017; Bau & Das, 2017). Consequently, there has been an increasing focus on teacher professional development programs.

There are a number of different models for teacher professional development programs. Some are focused on improving content and pedagogical content knowledge, while others are geared towards improving both simultaneously (Darling-Hammond et al., 2017). There is a lot of debate around which model works best. Research shows that the effectiveness of different models of training programs varies significantly across

contexts and applications, often determined by the grade and subject taught by the teacher (Hallman-Thrasher et al., 2019; Kraft, Blazar, & Hogam, 2017; Conn, 2014). However, to better understand which model works best in what context, we must understand the mechanisms or pathways through which the training programs are effective.

This study used longitudinal data on 444 teachers and 3,435 students to examine mechanisms that mediate the positive effect of a teacher professional development program on classroom quality and student outcomes in Ghana. Specifically, it examined whether teachers' knowledge of the learning content, teachers' implementation quality of behavioral and instructional practices and teachers' professional well-being were significant mediators of the treatment effect. It utilized a causal mediation approach, which allowed the average causal mediation effects to be parametrically and nonparametrically identified under a set of minimum conditions. The study found that teachers' implementation quality was a significant mediator of positive treatment effect on classroom quality across time. This effect persisted even in the presence of teacher knowledge and professional well-being. The study also found small marginal mediation effects on student outcomes, including a positive mediation effect on literacy and a negative mediation effect on executive functioning in the presence of all mediators. Overall, this study provides empirical evidence to design future interventions that place more emphasis on the influential pathway of implementation quality to yield positive impacts, particularly in early education contexts.

The second part of the study examined teacher profiles that provide diagnostic information about teachers' instructional strengths and weaknesses as measured by

classroom observation tools. Traditionally, researchers have employed a range of classroom observation tools to measure teacher quality and practices using frequency counts, time allocation and classroom management behaviors. However, scores from these instruments are often presented as an aggregate at the construct level or in some cases, at the item level. This is in contrast to the notion of teacher competencies framework (Blömeke & Delaney, 2012), and other studies that have found that effective teaching is in fact a function of multiple constructs related to cognitive abilities and affective-motivational characteristics that influences classroom quality and student outcomes (Kraft et al., 2017; Conn, 2014). To better understand teacher practices in a holistic manner and provide related feedback and support, it is perhaps more reasonable to understand how teachers perform across constructs rather than within each construct. Therefore, in this study, stage-wise cluster analysis was applied using constructs of teachers' professional well-being and classroom quality to group teachers into different subpopulations and study how they relate to student outcomes. The study found six profiles of teachers with varying professional well-being and classroom practices, including two that were significantly associated with positive student learning across all four domains of numeracy, literacy, socioemotional development and executive functioning. The first profile included teachers with average professional well-being and classroom quality while the second included teachers with good professional well-being and excellent classroom quality. This methodological approach allowed easy identification of instructional strengths and growth opportunities within each

subpopulation, which helps provide formative feedback and targeted support to facilitate high-quality teaching and maximize student learning.

While the first part of the study focuses on influential mechanisms that mediate positive impacts of comprehensive teacher professional development programs, the second part of the study hones in on understanding the performance of different subgroups of teachers by identifying their teaching needs across constructs that can be addressed with formative feedback and targeted support. Improved teacher performance across all domains of instructional practices is also likely to lead to pronounced positive impacts on student learning.

PART 1: WHAT MECHANISMS CAUSE TEACHER TRAINING PROGRAMS TO WORK?

1.1. Introduction

There is increasing evidence of a global "learning crisis" in international education. The World Development Report 2018 identifies three main dimensions of the learning crisis as follows: poor learning outcomes of children, lack of trained and motivated teachers and classroom facilities, and deeper systemic causes characterized by low accountability and high inequality (World Bank, 2018). Other research shows that the learning crisis is in fact a reflection of a teaching crisis (Bold et al, 2017). Studies in developing countries illustrate the need for high quality teachers to help students learn better. Buhl-Wiggers et al. (2017) find that a 1 SD increase in teacher quality can lead to up to .36 SD increase in student performance in reading in Uganda. Bau and Das (2017) find that a 1 SD increase in overall Teacher Value Added (TVA) can lead to a 0.21 SD increase in average test scores in Pakistan. Similarly, studies in United States also illustrate the importance of TVA over a single year and its relationship with short- and long-term student outcomes including achievement, college attendance, and labor market earnings (Chetty, Friedman and Rockoff, 2014; Kane and Staiger, 2008; Rockoff, 2004). Consequently, there has been a collective focus on improving teacher training programs around the world.

1.2. Teacher professional development programs

Currently, there exist many different models of teacher training programs. These range from content-focused programs to those that incorporate active learning, support

collaboration, use modeling techniques of effective practices, and provide ongoing support and feedback (Darling-Hammond et al., 2017). Some of these are short and last three to four days, while others are longer and of sustained duration. All these models are aligned with the tenets of adult learning theory (Swaner, 2016; Gregson & Sturko, 2007). However, there is a lot of debate around which model works best. There is no definite answer. Research shows that the effectiveness of different models of training programs varies significantly across contexts and applications. For example, content-focused teacher training programs tend to work well in improving instruction in core subject areas, such as mathematics, reading and science (Hallman-Thrasher et al., 2019; Garet et al., 2016; Thames, 2010). Lee (2017) found that this is true for teaching mathematics to children in preschool as well. However, other research suggests that programs that focus on delivery of actual teaching practices is very effective across subject areas in developed (Guskey, 2002; Guskey & Yoon, 2009; Kraft, Blazar, & Hogam, 2017) and developing countries (Abadzi, 2012; Conn, 2014). In a recent study, RTI International (2019) compared teacher training programs across seven countries and found that "not as yet successful" programs focus too much on content delivery, with little use of modeling techniques or active learning. Hamre et al. (2017) found that focus on actual delivery of instruction and classroom practices is particularly important in early childhood education and preschool settings.

1.3. Pathways of treatment effect

It is important to unpack the mechanisms at play that allow certain professional development programs to be more successful than others. Blömeke and Delaney (2012)

proposed a two-dimensional teachers' professional competencies framework that identified cognitive abilities and affective-motivational characteristics as the main levers of change. Cognitive abilities were further identified by content and pedagogical knowledge, while affective-motivational characteristics included professional beliefs about teaching and learning, motivation and self-regulation. I will review these dimensions in the light of literature from developed and developing countries.

Teachers' knowledge. There has been extensive research on teacher knowledge and how to conceptualize it. Shulman's (1986) work first defined the concept of pedagogical content knowledge. Many researchers iterated on the concept and described teacher knowledge as a combination of content and pedagogical content knowledge (Ball et al., 2005; Mishra, 2006; Burgess, 2009; Guerriero, 2014). Studies in developed countries show that better knowledge is indicative of higher student achievement; however, pedagogical knowledge may be more impactful (Guerriero, 2014). Similarly, studies in developing countries including Kenya show professional development programs can help improve teachers' literacy knowledge (Dubeck et al., 2015). However, follow-up studies have shown that literacy-based interventions, complemented with structured lesson plans and ongoing support to teachers, significantly improve classroom practices, student's literacy and retention (Jukes et al., 2016). Therefore, it appears that there are mechanisms beyond teacher knowledge that may also be crucial for positive student outcomes.

Teachers' instructional practices. Studies show that aspects of instructional practices are closely linked with teacher quality and student outcomes (Dobbie & Fryer,

2013; Hamre, 2014; Muijs et al., 2014). These aspects can be understood as characteristics of process quality and include well-structured classroom management, high quality teacher-student interactions, and cognitive activation with deep learning and critical thinking tasks (LoCasale-Crouch et al., 2016; Levya et al., 2015; Klieme et al, 2006). Meta-analyses show that successful in-service teacher training programs, complemented with coaching and support improve teacher's instructional practices and student outcomes in developed (Kraft, Blazar, & Hogam, 2018; Hattie, 2012) and developing countries (Ganimian & Murnane, 2016; Conn, 2017). In a meta-analysis of more than 1000 educational studies, Hattie (2012) found that teacher subject-matter knowledge had a smaller effect size (0.19 SD), compared to that of classroom management (0.52 SD) and effective teacher feedback (0.75 SD). Other evidence from United States also shows how effective coaching practices and teacher pedagogy can improve student outcomes (Teemant, 2014, Matsumura et al., 2010; Carlisle, Cortina & Katz, 2011). Similarly, Conn (2017) found that pedagogical interventions (changes in instructional techniques), especially interventions that use adaptive instruction and coaching techniques are particularly effective in developing countries, with a mean effect size of 0.23 SD on student outcomes. In another study, Piper et al. (2018) found that teacher instructional support and coaching was very successful in improving literacy and numeracy outcomes of first and second grade students in Kenya.

Teachers' well-being. Teacher well-being can be understood as a set of constructs that encompass cognitive, psychological, physical, and social well-being of teachers in schools (OECD, 2018). Teacher motivation is defined as intrinsic and extrinsic factors

that determine the value and hard work teachers put into teaching to ensure their students learn (Ryan & Deci, 2000). It is directly related to teacher quality and consequently a critical determinant of student outcomes (Collie & Martin, 2017; Klassen et al., 2012; Martin, 2009). Teacher burnout is understood as a prolonged emotional and interpersonal response to exhaustion, cynicism and inefficacy at work that critically impacts teaching outcomes (Maslach et al., 2001). It is linked with poor teacher quality (Greenberg et al., 2016; NASUWT, 2017) and low student achievement (McLean and Connor, 2015). In more extreme cases of demotivation and burnout, teachers leave their jobs or even the profession (Das et al., 2007; Mulkeen, 2010). This means that fewer teachers are available to support student learning and limited public sector resources spent on teacher training are wasted. Similarly, teacher job-satisfaction and personal accomplishment are often explained as dimensions of occupational well-being that are also important factors of teacher retention (Klusmann et al., 2008). Therefore, teacher well-being is a critical aspect of teachers' lives that must be taken into account to promote success in schools.

1.4. Context of Ghana

1.4.1. Overview

Ghana is a lower-middle-income country in West Africa and ranks 140 out of 189 countries on the Human Development Index – a composite index measure of life expectancy, education, and per capita income in a given country (UNDP, 2018). The Ministry of Education (MoE) is responsible for Ghana's education policies and systems, and the Ghana Education Service (GES) is the mandated agency for implementing interventions at the pre-tertiary level. Pre-tertiary education consists of basic and

secondary education. Basic education includes two years of Kindergarten (KG; KG1 – Kindergarten 1, which is equivalent to pre-K in the United States and KG2 – Kindergarten 2, which is equivalent to kindergarten in the United States), six years of Primary and three years of Junior High School (JHS). Secondary education consists of either three years of Senior High School (SHS) or technical/vocational education. According to a recent report by the World Bank, Ghana has highly prioritized the education sector, with total education expenditure exceeding international benchmarks. Between 2011 and 2015, education expenditure accounted for between 6 and 8 percent of country's GDP and 21 to 28 percent of government expenditure (Mikesell, 2018).

However, Ghana still faces some key challenges in education, related to student learning outcomes. The average years of schooling in Ghana are 11.6 but the number of quality adjusted learning years is only 5.7. This implies that children are attending school but learning at a very slow rate (Mikesell, 2018). The report identified low learning levels due to a host of reasons including poor infrastructure, ineffective teacher management, inefficient use of non-salary budget and inadequate accountability. Specifically, ineffective teacher management translated into inefficient teacher deployment, training and support to deliver curricula, and teacher absenteeism and attrition (Mikesell, 2018).

1.4.2. Educational teacher-related policies and practices

The Ministry of Education of Ghana recognizes teaching as a key driver of student learning and launched the Pre-Tertiary Teacher Professional Development and Management Policy to restructure professional development of teachers within appropriate competency frameworks in the pre-tertiary sector (Ministry of Education

Ghana Education Service, 2012). As a result, the government has invested in a multitude of pre-service and in-service teacher preparation programs to recruit and retain teachers in the education system. Traditionally, pre-service training has consisted of two years of coursework in a public college of education, followed by one year of fieldwork as a student-teacher, and then placement in a public school as a newly-qualified teacher (NQT). However, initiatives such as Transforming Teacher Education and Learning (T-TEL) are improving pre-service teacher preparation through public colleges of education in a recent initiative from 2014-2018 in Ghana (Cambridge Education, n.d.). T-TEL aimed to revamp pre-service teacher preparation by moving towards a five-year training model, in which students completed a four-year bachelor's degree followed by one year of teaching in a basic school, in order to get a license to practice and achieve a qualified teacher status (Kale-Dery, 2018).

1.5. Research questions

There have been relatively few randomized evaluations of teacher training programs in Ghana. Most recently, Wolf (2018) found that a pre-service teacher training program in rural Ghana improved teacher knowledge and implementation of national curriculum, but there were mixed impacts on professional well-being and no impacts on student learning. However, in another evaluation, Wolf et al. (2018a) found that a short, in-service teacher training program improved teaching, classroom quality, and school readiness of kindergarten children, and some of the impacts persisted in the following school year.

This study builds on this previous evaluation by Wolf et al. (2018a). The teacher training program was implemented in six districts of the Greater Accra Region across public and private schools serving children (aged 4-6 years) enrolled in kindergarten. The researchers randomly assigned 240 schools to either receive the a) teacher training/coaching, b) teacher training/coaching along with a parental awareness program, or c) neither (comparison group). 82 schools belonged to the first treatment arm, 79 belonged to the second treatment arm while another 79 belonged to the comparison group. The training began with a five-day course, followed by a two-day refresher training four months later, and a one-day refresher four months after that. The program offered experiential training for teachers and included ongoing monitoring and feedback. It focused on helping teachers learn age-appropriate play-based instructional techniques that build a positive classroom environment. The evaluation revealed moderate positive impacts on dimensions of teachers' professional well-being and classroom quality, and small impacts on multiple domains of students' school readiness outcomes, with some of the impacts persisting in the following school year.

Given the positive impacts of the intervention, this study aims to understand the potential mechanisms that mediate the causal effect of teacher training on classroom quality and student outcomes. The research questions are as follows:

1. Is the treatment effect on classroom quality and student outcomes mediated by teachers' content knowledge, teachers' implementation quality of teaching practices, and/or teachers' professional well-being?
2. If so, are these independent (simple) or related (multiple) mediation effects?

3. Are these mediation effects sustained over time?

The study makes two important contributions. The first contribution is methodological. This study utilizes a causal mediation framework and assess the validity of the estimates through a set of sensitivity analyses. To the best of my knowledge, this method has not been used to understand the causal pathways through which teacher training programs have an impact. The second contribution is empirical. The study provides empirical estimates of the contribution of different mechanisms that are hypothesized to impact the effectiveness of teacher training programs. These help us understand why some mechanisms work better than others and why there may be a need to design future interventions that place more emphasis on the influential pathways to yield positive impacts, particularly in early education contexts.

1.6. Sample

The teacher training program was implemented in the six of the nine most disadvantaged districts (according to 2014 UNICEF District League Table) in the Greater Accra region. These included Ga South, Adenta, Ledzokuku-Krowor, Ga Central, La Nkwantanang-Madina, and Ga West.

In these six districts, 240 schools were identified using the GES Educational Management Information System (EMIS) database. Of the 240 schools, 108 were public and 132 were private. All KG teachers in these schools were invited to participate in the training program. If there were more than two KG teachers in each school, two were randomly selected for the evaluation (one from KG1 and one from KG2). However, 36 schools had only one KG teacher. The final sample included 444 teachers. 98% of the

teachers were female. The mean age at baseline was 35; the youngest was 18 while the oldest teacher was 69.

Additionally, 15 students were randomly selected for direct assessments from each school roster (eight from KG1 and seven from KG2). Assessors also randomly selected up to 10 additional children on the initial visit for a reserve list. If a selected child from the first 15 was not in the school that day, assessors returned up to two times to assess the child. If the child was still not present on the third day visit, a child from the reserve list was selected instead. If a school had fewer than 15 KG children, then all children were selected. The final sample included 3,435 children at baseline. 51% of students were male while remaining were female. The mean age at baseline was 5.2 years; the youngest was 3 while the oldest child was 11.

Refer to Wolf et al. (2018a) for a detailed description of the randomization and sampling procedures.

1.7. Procedures

The data were collected in three rounds over a period of two academic school years: baseline, follow-up I and follow-up II. Baseline data were collected at the start of the academic year in September-October 2015. Follow-up I data were collected at the end of the academic year in May-June 2016. Follow-up II data were collected in the following academic year in May-June 2017. The data was collected using teacher surveys, classroom observations and child assessments. For classroom observations, teachers were videotaped teaching a lesson in their classrooms for 30 to 60 minutes. The

videos were then coded using an implementation fidelity checklist and a tool to assess classroom quality.

There was some attrition at school, teacher and student level. Of the 240 schools sampled at baseline, only five schools dropped out at follow-up I and follow-up II. Three of these declined to participate in the study while two were closed down due to land litigation issues. Of the 444 teachers selected at baseline, 348 teachers were interviewed at follow-up I but three of the teachers were missing data on classroom observations, resulting in an analytic sample of 345 teachers at follow-up I. Of these 345, 309 teachers were interviewed at follow-up II but again, fourteen teachers were missing data on classroom observations, resulting in an analytic sample of 295 teachers at follow-up II. Most teachers dropped out either due to the transfer of teachers from pre-school classes or teachers leaving the profession. Finally, of the 3435 children assessed at baseline, there were 2358 children for the 345 teachers at follow-up I and 1883 children for 295 teachers at follow-up II. Most children dropped out due to the change in schools because of family migration out of the district or region.

1.8.Measures

1.8.1. Mechanisms

Knowledge. Teacher knowledge was assessed using an aggregate measure of self-reported perceptions of early childhood development in three domains: developmentally appropriate practice, social-emotional development and family-supportive practice. The first two can be understood as dimensions of content and pedagogical knowledge (Guerriero, 2014). However, family-sensitive caregiving is an additional construct

considered very relevant in early education settings (Connor et al., 2005). Note teacher knowledge was not measured at baseline. The resulting aggregate score demonstrates adequate internal consistency (Follow-up I: M =4.52, SD =.34, α =.79; Follow-up II: M =4.52, SD =.32, α =.77).

Developmentally appropriate practice was measured using six items on a scale of 1 (not very important) to 5 (very important). Items assessed teachers' perceptions of children's needs as they "grow and develop", encourage them to recognize "letters or words" and "numbers or shapes", to work with families to set "individual plans and goals for children", provide "materials for play and learning" and measure children's development over time to determine how they're doing (Follow-up I: M =4.75, SD =.30, α =.63; Follow-up II: M = 4.73, SD =.27, α =.49).

Supporting children's social and emotional development was measured using five items on a scale of 1 (not very important) to 5 (very important). Items assessed teachers' perceptions of children's needs to "build relationships with peers and adults", "learn to control their behavior", "express thoughts and feelings", "resolve conflicts with other children", and have behavior guidance styles that "match the parents" (Follow-up I: M =4.64, SD =.39, α =.63; Follow-up II: M = 4.63, SD =.35, α =.53).

Family-sensitive caregiving was measured using five items on a scale of 1 (not very important) to 5 (very important). Items assessed teachers' perceptions on considering "parents' goals, ideas and suggestions", willing to "work with parents about their work schedules", including families in "decision-making for child's education", caring about the "entire family, not just the child", and connecting "families to outside or

community resources" (Follow-up I: M =4.14, SD =.62, α =0.71; Follow-up II: M = 4.17, SD =.60, α =.71).

Implementation. Teacher implementation quality was assessed using a checklist of 15 activities that were explicitly covered in the teacher training related to teaching practices. Of the 15 items, 5 items measured behavioral practices while the remaining measured instructional practices. Items measuring behavioral practices included checks such as "teacher praises children for positive behavior", "teacher explicitly reminds children of the class rules", or "teacher uses a signal to gain children's attention". Items measuring instructional practices included checks such as "teacher reads a storybook during the lesson", "there is an activity that facilitated the lesson objectives that involved manipulation of materials", or "teacher asks students at least two open-ended questions during the class period". Each practice was coded as either present (=1) or absent (=0) in recorded videos of teachers teaching a lesson (Baseline: M = 2.23, SD = 1.85; Follow-up I: M = 4.32, SD = 1.89; Follow-up II: M = 3.46, SD = 1.59).

***Well-being*.** Teacher professional well-being was measured using self-reported measures of motivation, job satisfaction and burnout.

Motivation was measured using a scale originally developed by Bennell and Akyeampong (2007) as reported in Wolf et al. (2015). Items assessed teachers' motivation to help children learn to "read and write", "develop well emotionally", "develop well socially" and overall personal attitude towards teaching. The scale consists of five items, with scores ranging from 1 to 5 (1=false, 2=mostly false, 3=sometimes true, 4=mostly true, 5=true) for each item. The resulting average score at follow-up I and II

demonstrates adequate internal consistency (Baseline: $M = 4.66$, $SD = .58$, $\alpha = .57$; Follow-up I: $M = 4.71$, $SD = .44$, $\alpha = .77$; Follow-up II: $M = 4.63$, $SD = .47$, $\alpha = .69$).

Burnout was measured using 11 items from the Maslach Burnout Inventory, scored on a

scale of 1 (never) to 7 (everyday) to indicate the level of fatigue from work (Maslach et al.,

1996). Items assessed teacher's emotional exhaustion form work, fatigue at the start and end of the day and general attitude towards peers and children. The resulting average score demonstrates adequate internal consistency (Baseline: $M = 2.06$, $SD = .92$, $\alpha = .69$; Follow-up I: $M = 2.01$, $SD = .89$, $\alpha = .75$; Follow-up II: M = 2.09, SD = .91, $\alpha = .78$).

Job satisfaction was measured using another scale developed by Bennell and Akyeampong (2007) and used in Wolf et al. (2015). The scale consists of six items, with scores ranging from 1 to 4 (1 = true, 2 = somewhat true, 3 = somewhat false, 4 = false; items were recoded so higher scores reflect higher job satisfaction). Items assessed overall satisfaction with the job at the school, decision to be a teacher and commitment to the teaching profession. The resulting average score demonstrates adequate internal consistency (Baseline: $M = 1.84$, $SD = .63$, $\alpha = .64$; Follow-up I: $M = 3.09$, $SD = .68$, $\alpha = .73$; Follow-up II: $M = 3.04$, $SD = .68$, $\alpha = .75$).

1.8.2. Outcomes

Classroom quality. Classroom quality was measured as an aggregate of three domains of teacher–child interactions found in Wolf et al. (2018b): facilitating deeper learning, supporting student expression and emotional support and behavior management.

Each of these domains was measured using the Teacher Instructional Practices and Processes System (TIPPS; Seidman et al., 2018; Seidman et al., 2014). The TIPPS is a classroom observation tool for assessing classroom quality focused on teacher-child interactions. The resulting aggregate score demonstrates adequate internal consistency (Baseline: M =2.13, SD =.29, α =.80; Follow-up I: M =2.33, SD =.25, α =.75; Follow-up II: M =2.35, SD =.30, α =.76).

Facilitating deeper learning was measured using three items on whether teacher connects lesson to teaching objectives, scaffolding and high-quality feedback (Baseline: M =2.03, SD =.69, α =.55; Follow-up I: M =2.31, SD =.58, α =.42; Follow-up II: M =2.37, SD =.74, α =.50).

Emotional support and behavior management was measured using seven items on positive climate, teacher sensitivity and responsiveness to student needs, and providing consistent routines (Baseline: M =2.77, SD = .45, α =.73; Follow-up I: M =3.07, SD = .37, α =.83; Follow-up II: M =2.93, SD =.37, α =.81).

Supporting student expression was measured using four items related to whether teacher considers student ideas during the lesson and encourages students to reason and problem solve (Baseline: M =1.48, SD =.52, α =.57; Follow-up I: M =1.65, SD =.55, α =.63; Follow-up II: M =1.86, SD =.64, α =.61).

Student learning. Student outcomes were assessed in four domains of school-readiness including literacy, numeracy, socioemotional and executive functioning. Each of these domains was measured using the International Development and Early Learning Assessment tool developed by Save the Children (Pisani et al., 2015).

Early literacy was measured using a scale of 38 items grouped into six subtasks related to print awareness, letter knowledge, phonological awareness, oral comprehension, emergent writing, and expressive vocabulary (Baseline: M =.46, SD =.22, α = .74; Follow-up I: M =.61, SD =.20, α = .72; Follow-up II: M =.70, SD =.18, α = .88).

Early numeracy was measured using a scale of 39 items grouped into eight subtasks related to 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 (Baseline: M =.44 , SD =.19 , α = .72; Follow-up I: M =.57 , SD =.19 , α = .70; Follow-up II: M =.67, SD =.16 , α = .72).

Social-emotional development was measured using a scale of 14 items grouped into five subtasks related to self-awareness, emotion identification, perspective taking and empathy, friendship, and conflict and problem solving (Baseline: M =.41, SD =.20 , α = 0.69; Follow-up I: M =.54, SD =.19, α = .70; Follow-up II: M =.58, SD =. 17, α = .67).

Executive functioning was measured using a scale of ten items grouped into two subtasks related to working memory and impulse control (Baseline: M =.49, SD =.21 , α = .84; Follow-up I: M =.59, SD =.18, α = .83; Follow-up II: M =.64, SD =.16, α = .79).

1.8.3. Covariates

A small set of covariates was used, including teacher gender, age in years, level of education, years of teaching experience, a baseline score for each respective outcome to control for teacher related heterogeneity. Similarly, child gender, age in years, KG level, and baseline score for each respective outcome were included to control for student

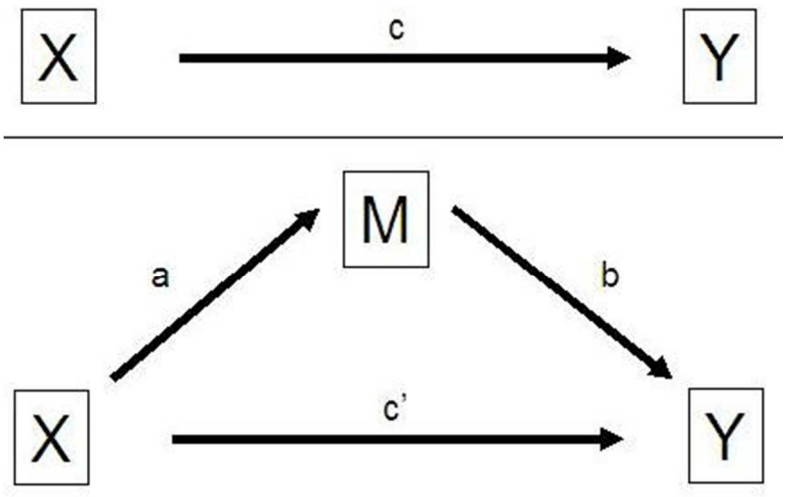
related heterogeneity. Due to the randomization of the treatment, the treatment and control groups were expected to be similar, on average. While the unbiased estimation of the treatment effect did not require the inclusion of additional covariates, the causal mediation analysis imposes additional assumptions (described later) that are likely to be more plausible with the inclusion of additional covariates. Additionally, inclusion of covariates also improves the precision of the estimates.

1.8. Methods

1.8.1. Earlier approaches

Within the field of statistics, numerous approaches have been developed to perform mediation analyses. Some of these include the causal steps strategy proposed by Barron and Kenny (1986), product of coefficients approach or also known as Sobel test (Sobel, 1982; 1986), and bootstrapping procedure introduced by Preacher and Hayes (2008). Figure 1 shows the general framework used in mediation analyses, where X is the exogenous treatment, M is the potential mediator and Y is the outcome. The total effect of X on Y is captured by c, while the direct effect of X on Y, in the presence of a mediator is captured by c'.

Figure 1: Causal mediation framework illustrating total effect of X on Y, direct effect and indirect effect through M.



The causal steps strategy assesses the extent to which several criteria are met. These include whether X is related to M, X is related to Y, and M is related to Y. If these criteria are met, then the association between X and Y decreases substantially when M is added as a predictor. The product of coefficients approach or Sobel test does not focus on individual paths but on the product term ab, under the logic that it is equal to the difference between the total and direct effect (c-c'). It assumes the difference to be normally distributed for inferences. However, this is only true in the case of very large samples. The bootstrapping procedure is a non-parametric resampling procedure that does not assume the difference to be normally distributed. The procedure yields percentile bootstrap confidence intervals, which unlike regular confidence intervals, can be asymmetrical since they are based on an empirical estimation of the sampling distribution. This is the approach used in this study.

1.8.2. Causal mediation approach

The causal mediation approach builds upon these approaches and has the potential to address the common criticism of experiments that they only present a "black-box view of causality" (Imai et al., 2011). It assumes that under a set of minimum conditions, the product of coefficient method and its variants yield valid estimates of the average causal mediation effect (ACME) (Imai et al., 2010; Tingley et al., 2014). The minimum conditions are captured by the sequential ignorability assumption, which states that the observed mediator status is as if randomly assigned conditional on the randomized treatment variable and the pretreatment covariates. Therefore, under this assumption, causal mediation analysis requires two statistical models: one for the mediator $f(M_i \mid T_i, X_i)$ and the other for the outcome variable $f(Y_i \mid T_i, M_i, X_i)$ where T_i is the treatment status for teacher i, M_i is the mediator and X_i is a set of pretreatment covariates. Once these models are chosen and estimated, the causal mediation and other relevant estimates will be computed using the algorithms proposed in Imai et al. (2010). The algorithms also produce confidence intervals based on either a non-parametric bootstrap procedure (for parametric or nonparametric models) or a quasi-Bayesian Monte Carlo approximation (for parametric models).

1.8.3. Sensitivity analysis

An important contribution of this approach is a set of sensitivity analyses that can be performed to formally quantify the robustness of the estimates of the mediation effect to the potential violation of sequential ignorability assumption – which is the "key and yet untestable assumption for identification" (Imai et al., 2010). The challenge is that there may exist unobserved confounders that causally affect both the mediator and the

outcome even after conditioning on the observed treatment and pretreatment covariates. Therefore, assessing the sensitivity of the empirical findings to the possible existence of such confounders is required in order to evaluate the validity of the mediation study.

Sensitivity analysis provides two sensitivity parameters. The first sensitivity parameter is the correlation ρ between the residuals of the mediator and outcome regressions. If unobserved confounder affecting the mediator and outcomes exist, then ρ is no longer 0 and the sequential ignorability assumption is violated. The second sensitivity parameter, which is mathematically equivalent to ρ, is the product of R^2 (or coefficient of determination) of the two regressions. This suggests that the omitted confounders have to explain a proportion of variation to invalidate the estimated ACME values.

1.8.4. Multiple mediation

The causal mediation approach has been extended to accommodate multiple mediators by generalizing a linear structural equation model in several ways. First, the model permits the presence of causally dependent multiple mediators. Second, it allows the coefficient estimates to vary across individual observations for heterogenous treatment effects. Third, it includes interactions between the treatment and each of the mediators so that the mediation effects can vary by the baseline treatment status. However, when posttreatment confounders are causally related (or equivalently, when one mediator acts as a posttreatment confounder for the other mediator on the outcome), Imai and Yamamoto (2013) propose that the ACMEs can be estimated under the simple mediation model with sequential ignorability assumption if an additional assumption of

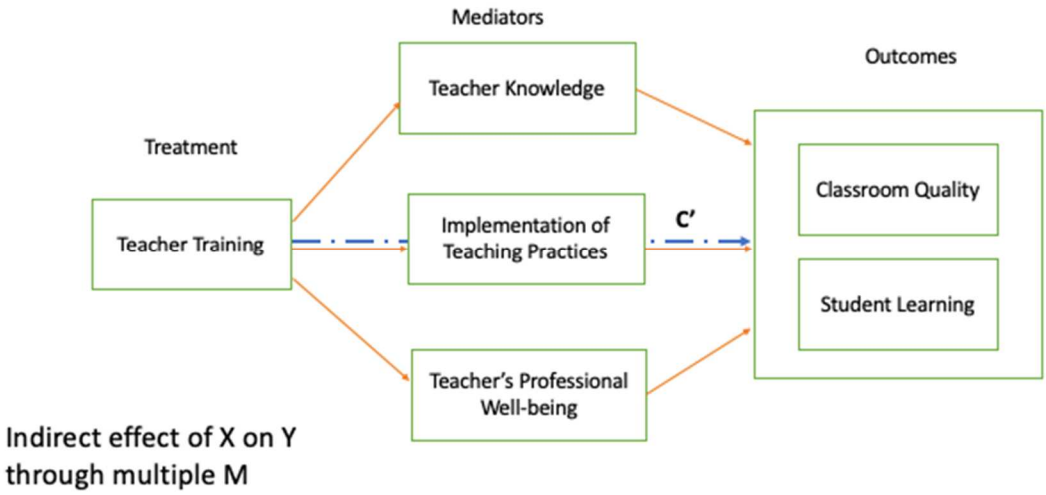
homogenous interaction is satisfied (Imai and Yamamoto, 2013; Tingley et al., 2014). This additional assumption states that the degree of interaction between the treatment and the primary mediator is constant across individual units. However, if the additional assumption is violated, sharp bounds on the ACME as functions of a parameter representing the degree of the violation can be expressed as part of the sensitivity analysis. The sensitivity parameter is the standard deviation of the coefficient on the treatment-mediator interaction term. Two alternative sensitivity parameters based on coefficient of determination include the proportion of the residual or the total variance of the outcome variable explained by including the heterogeneity in the treatment-mediator interaction in the model. These parameters represent how important it is to incorporate the interaction heterogeneity to explain variation in the outcome.

1.8.5. Packages in R

The causal mediation analysis for this study was conducted in R, using the *mediation* package that allows tests for simple mediation (*mediate*), multiple mediation (*multimed*) and sensitivity (*medsens*) of the estimates (Tingley et al., 2014). To implement the Barron-Kenny procedure in *mediation*, linear models for both the mediator and outcome were estimated, controlling for school (private/public ownership of the school), teacher (gender, age, experience) and student (gender, age) characteristics. The model objects from these two parametric models were input into the *mediate* function, which estimated the ACMEs with non-parametric bootstrap confidence intervals. The analyses were followed with a set of sensitivity analyses and presented estimates of sensitivity parameters where ACMEs would equal zero.

For multiple mediation, a data frame containing the necessary variables (outcome, primary mediator, alternative mediator, treatment and pre-treatment covariates) was input into the *multimed* function, which again estimated the ACMEs with non-parametric bootstrap confidence intervals under the sequential ignorability and homogenous interaction assumptions. It presented three variants of the ACME, including the ACME for the overall sample, and the treatment and control group, respectively to allow for the possibility that ACME may differ depending on the baseline treatment status. Additionally, the *multimed* function conducted a sensitivity analysis with respect to possible heterogeneity in the treatment-mediator interactions. Similar to simple mediation, it presented values of sensitivity parameters where the estimated ACMEs would equal zero. Figure 2 presents the causal mediation framework for this study.

Figure 2: Causal mediation framework for this study.



1.8.6. Missing data

Earlier analyses by Wolf et al. (2018a) show that teachers who dropped out at follow-up I were not systematically different from the ones who remained in the sample. Consequently, the results are not biased by excluding them from the analysis. However, there was still some missing data for a subset of remaining teachers at follow-up I and follow-up II, respectively. Multiple imputation was used (with the *mice* package in R) to address missing data for teacher and classroom characteristics. A rich set of covariates from baseline and follow-up I (including teacher demographic and background variables, outcome scores for teacher knowledge, implementation and classroom quality) was used to impute teacher-level missing data for the 345 teachers at follow-up I. Similarly, the same set of covariates from baseline, follow-up I and follow-up II was used to impute teacher-level missing data for the 295 teachers at follow-up II. Tables A1-A3 (Appendix A) present the descriptive statistics for teacher-level data for the raw and imputed datasets, respectively. The statistics are largely the same for all variables across the datasets.

Data was not imputed for students due to a sufficiently large sample size at follow-up I and follow-up II, respectively. Table A4 (Appendix A) presents the descriptive statistics of the final analytic sample of students at the three rounds of data collection.

1.9. Results

1.9.1. Mediation effects: Classroom quality

***Simple Mediation*.** Table 1 presents estimates for the ACME, average direct effect (ADE) and total effect for each of the mediators on classroom quality at follow-up

I and follow-up II, respectively. Controlling for school characteristics (private/public ownership of the school) and teacher characteristics (gender, age, experience), teacher knowledge did not mediate the effect of teacher training on classroom quality. However, a SD increase in the implementation quality of the teaching practices positively mediated the treatment effect by .34 SD ($p < .01$) and offset the negative ADE of -.04 SD ($p > .1$) to yield a total effect of .30 SD ($p < .1$). Therefore, the implementation quality of teaching practices mediated the total effect by 112%. At follow-up II, a SD increase in the implementation quality of the teaching practices positively mediated the treatment effect by a smaller magnitude of .17 SD ($p < .01$) and partially offset the ADE of -.36 SD ($p < .05$) to yield a total effect of -.19 SD ($p > .1$). Professional well-being of teachers, including motivation, job satisfaction and burnout did not significantly mediate the effect of teacher training on classroom quality in follow-up I or follow-up II.

Table 1: Estimates for simple mediation on classroom quality.

	Follow-up I (June 2016)					Follow-up II (June 2017)				
	Estimate	95% CI Lower	95% CI Upper	p-value		Estimate	95% CI Lower	95% CI Upper	p-value	
Knowledge										
ACME	-0.02	-0.06	0.02	0.46		-0.01	-0.05	0.03	0.78	
ADE	0.32	0.10	0.56	0.00	**	-0.19	-0.47	0.09	0.19	
Total Effect	0.30	0.08	0.55	0.00	**	-0.19	-0.47	0.09	0.17	
Prop. Mediated	-0.05	-0.30	0.09	0.46		0.03	-0.42	0.59	0.82	
Implementation										
ACME	0.34	0.22	0.48	0.00	***	0.17	0.08	0.28	0.00	***
ADE	-0.04	-0.29	0.20	0.78		-0.36	-0.63	-0.10	0.00	**
Total Effect	0.30	0.05	0.54	0.01	*	-0.19	-0.45	0.07	0.14	
Prop. Mediated	1.12	0.55	4.58	0.01	*	-0.88	-5.82	5.29	0.14	
Motivation										
ACME	-0.01	-0.04	0.02	0.51		0.01	-0.01	0.05	0.45	
ADE	0.31	0.08	0.55	0.01	**	-0.20	-0.48	0.05	0.13	
Total Effect	0.30	0.06	0.54	0.01	**	-0.19	-0.47	0.07	0.15	

Prop. Mediated	-0.03	-0.27	0.06	0.52		-0.05	-0.85	0.52	0.56
Job Satisfaction									
ACME	0.00	-0.01	0.03	0.72		0.00	-0.02	0.01	0.92
ADE	0.30	0.06	0.53	0.02	*	-0.19	-0.45	0.06	0.15
Total Effect	0.30	0.06	0.53	0.02	*	-0.19	-0.45	0.06	0.14
Prop. Mediated	0.01	-0.07	0.12	0.72		0.00	-0.15	0.26	0.94
Burnout									
ACME	0.01	-0.03	0.05	0.57		0.00	-0.03	0.01	0.81
ADE	0.29	0.07	0.53	0.01	*	-0.19	-0.46	0.07	0.16
Total Effect	0.30	0.08	0.53	0.01	*	-0.19	-0.46	0.07	0.16
Prop. Mediated	0.04	-0.14	0.21	0.57		0.01	-0.18	0.29	0.83

Note. Estimates under the sequential ignorability assumption.

Sensitivity analysis. To test the robustness of these results, sensitivity analyses proposed by Imai et al. (2010) were conducted. As discussed earlier, sequential ignorability is an important and untestable assumption. The sensitivity analysis quantifies the extent to which the omitted confounder would need to be correlated with both the mediator and the outcome after conditioning on the observed treatment and pretreatment covariates to invalidate the mediation effects. The analysis shows that if the correlation between the error terms in the mediator and the outcome models (ρ) was .30 at follow-up I, the ACME of implementation quality of teaching practices may be 0. This implies that $\rho >$.30 will invalidate the findings. Other studies within political psychology also consider results to be highly robust when $\rho \leq .43$, see Imai & Yamamoto (2013). However, it may be easier to interpret the sensitivity of the findings in terms of the coefficient of determination parameters. In this case, if the product of R_M^2 and R_Y^2 was .09, the omitted confounders must explain 25% of the variation in the implementation quality of the teaching practices and 35% of variation in classroom quality, for example.

Similarly, ρ must equal .18 or product of R_M^2 and R_Y^2 must be .04 (i.e. omitted confounders must explain 20% of the variation in implementation and classroom quality, respectively) at follow-up II to invalidate the findings. Impact studies of teacher professional development programs within sub-Saharan Africa have found that all baseline covariates tend to explain 15-30% of the variation in teacher and student outcomes (Cilliers et al., 2019; Piper et al., 2018; Kelcey et al., 2016). Given this, it is unlikely that omitted confounders alone will explain 20-35% of the variation in classroom quality. Overall, the results show that under the sequential ignorability assumption, the ACME of implementation quality of teaching practices was positive and statistically significant, and the estimates were fairly robust to the possible presence of unobserved pretreatment mediator-outcome confounding. Thus, the implementation quality of teaching practices positively mediated the classroom quality. Figure B1 (Appendix B) presents the ACME, ADE and total effect of implementation quality of teaching practices on classroom quality, along with the sensitivity of the mediation effect.

Multiple Mediation. Table 2 presents the estimated ACMEs (conditional on treated, control and weighted average of treated and control) of implementation quality of teaching practices on classroom quality, accounting for the relationship with other mediators including teacher knowledge and professional well-being under the homogenous interaction assumption. The overall ACME was .33 SD ($p < .05$) at follow-up I and .19 SD ($p < .05$) at follow-up II. Compared to simple mediation effect of implementation quality, the multiple mediation effects were less statistically significant.

Table 2: Estimates for multiple mediation on classroom quality.

	Follow-up I (June 2016)			Follow-up II (June 2017)		
	Estimate	95% CI Lower	95% CI Upper	Estimate	95% CI Lower	95% CI Upper
ACME (treated)	0.36	0.23	0.50	0.18	0.06	0.30
ACME (control)	0.24	-0.03	0.52	0.22	0.02	0.42
ACME (average)	**0.33**	**0.20**	**0.47**	**0.19**	**0.07**	**0.30**
ADE (treated)	0.06	-0.27	0.39	-0.34	-0.66	-0.01
ADE (control)	-0.06	-0.31	0.19	-0.29	-0.57	-0.02
ADE (average)	0.03	-0.26	0.31	-0.33	-0.62	-0.03
Total Effect	0.30	0.08	0.51	-0.12	-0.38	0.16

Note. Estimates under the sequential ignorability and homogeneous interaction assumptions.

Results from the sensitivity analyses also showed that under the sequential ignorability and homogenous interaction assumptions, the ACMEs were moderately sensitive to the interaction heterogeneity. This was indicated by the confidence intervals which began to cover 0 as the homogenous interaction assumption was relaxed and the lower bound became less than 0 when σ > .40 at follow-up I and σ > .23 at follow-up II. Similarly, in terms of R^2 parameter, the value of .10 at follow-up I and .04 at follow-up II implied that the interaction heterogeneity must explain 10% of the total variation in classroom quality at follow-up 1 and 4% of the total variation in classroom quality at follow-up II so that the bounds of the ACME contain 0. Nevertheless, this does not invalidate the mediation effects, but implies that the positive treatment effects on classroom quality were mediated by heterogenous treatment-mediator interactions. Figure B2 (Appendix B) presents the ACME, ADE and total effect for multiple mediation on classroom quality, along with the sensitivity of the mediation effects.

1.9.2. Mediation effects: Student Outcomes

Simple Mediation. Tables C1-C4 (Appendix C) present estimates for the ACME, ADE and total effect for each of the mediators on student learning (numeracy, literacy,

socioemotional development and executive functioning) at follow-up I and follow-up II, respectively. Both linear models for the mediator and outcome controlled for school (private/public ownership of the school), teacher (gender, age, experience) and student (gender, age) characteristics. The magnitude of the ACMEs was very small and marginally significant across student outcomes.

Numeracy. At follow-up I, teacher knowledge and burnout positively mediated the treatment effect on numeracy by .01 SD ($p < .1$). At follow-up II, the implementation quality of teaching practices positively mediated the treatment effect by .02 SD ($p < .1$) and teacher motivation by .01 SD ($p < .05$).

Literacy. The implementation quality of teaching practices positively mediated the treatment effect on literacy by .03 SD ($p < .1$) at follow-up I and by .02 SD ($p < .1$) at follow-up II.

Socioemotional development. There were no mediation effects on socioemotional development at follow-up I. However, teacher motivation positively mediated the treatment effect by .01 SD ($p < .05$) at follow-up II.

Executive Functioning. At follow-up I, the implementation quality of teaching practices negatively mediated the treatment effect on executive functioning by -.07 SD, while motivation positively mediated by .01 SD ($p < .1$) and burnout by .02 SD ($p < .05$). There were no mediation effects on executive functioning at follow-up II.

Due to the marginally significant and extremely small mediation effects on student outcomes, the sensitivity analyses showed that the results were very sensitive to potential presence of omitted confounders i.e. any omitted confounder explaining a very

small proportion of variance (close to null) in the outcome variable can invalidate the findings. These may include parent-teacher relationships, parent-child interaction or family characteristics that may influence the mediators and student outcomes (Connor et al., 2005).

Multiple Mediation. Tables C5-C8 (Appendix C) present the estimated ACMEs of implementation quality of teaching practices on student outcomes, accounting for teacher knowledge and professional well-being under the homogenous interaction assumption. There were no significant mediation effects on student numeracy and socioemotional development. However, there was an overall positive mediation effect on student literacy ($ACME = .04, p < .05$), and surprisingly a negative mediation effect on executive functioning ($ACME = -.07, p < .05$) at follow-up I. There were no significant mediation effects on student outcomes at follow-up II.

Despite significant estimates, sensitivity analyses showed that the results were very sensitive to the heterogeneity in the treatment-mediator interactions. As was the case with classroom quality, the treatment effect on student literacy and executive functioning was mediated by heterogenous treatment-mediator interactions.

In summary, impacts on classroom quality were positively mediated by the implementation quality of the teaching practices, even when other potential mediators were allowed to be related. The mediation effects were moderate, robust to the presence of omitted confounders and persisted over time (Follow-up I = .34 SD, Follow-up II =.17 SD). Impacts on different student outcomes were mediated by a combination of mediators including implementation quality, teacher knowledge, motivation and burnout.

However, the mediation effects were very small and often marginally significant. Additionally, the mediation effects were also very sensitive to the presence of omitted confounders, therefore providing little conclusive evidence. Studies such as randomized controlled trials that are designed to specifically examine causal pathways can further test the robustness of these findings.

1.10. Discussion

With an increasing focus on delivering the "promise of education", practitioners and researchers world-wide are interested in understanding what type of educational interventions are particularly successful and examining the mechanisms that make certain interventions more successful than others to scale across contexts and countries. Amidst this focus, there has been increasing research on teacher training and professional development programs that have shown promising results for students' learning and development. These interventions have spanned across different models of pedagogical and instructional techniques. Some of these include programs focused on improving content delivery in core competencies while others have furthered methods for actual delivery of instructional practices with ongoing mentorship and feedback. Often, many of the trainings and professional development programs are a combination of both. However, little is known of what mechanisms play a role to allow certain training models to be more effective, and in which contexts the effectiveness of the program persists the most over time. To unpack this complexity, this study aimed to understand some of the potential mechanisms that allow teacher training programs to be effective, specifically in early learning contexts.

The study utilized a sample of 444 kindergarten teachers in the Greater Accra Region of Ghana who had been randomized to receive a teacher training/coaching intervention. Earlier research by Wolf et al. (2018a) found moderate positive treatment impacts of this intervention on some dimensions of professional well-being, classroom quality and small impacts on multiple domains of student outcomes. Some of these positive impacts persisted over time. Given the positive impacts, this study examined the causal mechanisms that mediated the treatment effect on classroom quality and student outcomes. Specifically, it examined the role of teacher knowledge, implementation quality of teaching practices and professional well-being in promoting a positive classroom environment and facilitating student learning.

1.10.1. Classroom Quality

Results for simple mediation showed that the implementation quality of the teaching practices was the only significant mediator of treatment effect on classroom quality with an ACME of .34 SD at follow-up I. The mediation effect persisted over time with an ACME of .17 SD at follow-up II. The results were fairly robust to the presence of unobserved omitted pretreatment confounders, even when other mediators including teacher knowledge and professional well-being were allowed to relate with implementation quality. However, the results were sensitive to treatment-interaction heterogeneity, which implied that the mediators heterogeneously influenced the treatment effect across individual observations. Nevertheless, implementation quality was the primary mediator of the treatment effect.

As a reminder, implementation quality was assessed using a checklist for a set of instructional and behavioral practices. Earlier studies have shown that the delivery of teaching practices is a key aspect of classroom quality that predicts student learning outcomes (Hamre et al., 2017; Hamre et al., 2014; Hamre, 2014; Pianta & Hamre, 2009) However, there is a growing interest in the implementation quality of teaching practices (Berkel et al., 2011; Durlak & DuPre, 2008), which entails an evaluation of whether the core intervention components have been delivered in a clear and comprehensible manner (Berkel et al., 2011). While implementation quality is well measured across a wide range of practices in this study, a more nuanced measure may show even greater mediation effects. Research shows that mean effect sizes can double or even triple if implementation with high fidelity is monitored (Durlak & DuPre, 2008). On the other hand, if implementation is not well-managed, negative impacts become evident (Hulleman & Cordray, 2009). Nevertheless, the way teachers implemented teaching practices, especially when they had been instructed to deliver the intervention in a standardized manner with ongoing provision of supports to maintain fidelity, was closely linked with classroom quality and student learning (Domitrovich & Greenberg, 2004). In some cases, teachers' perception of intervention quality also related to how they responded to interventions (LoCasale-Crouch et al., 2016).

Other research shows that teacher knowledge is also an important mechanism, but it has a much smaller effect size, compared to that of classroom management and effective teacher feedback (Hattie, 2012). Similarly, literacy-based interventions complemented

with structured lesson plans and ongoing support to facilitate implementation improve classroom quality and student learning (Jukes et al., 2017).

1.10.2. Student Outcomes

Results for simple mediation showed extremely small or negligible mediation effects on student outcomes. Specifically, teacher knowledge positively mediated the treatment effect on numeracy with an ACME of .01 SD at follow-up I. Implementation quality positively mediated the treatment effect on literacy with an ACME raging between .02 and .03 SD at follow-up I and follow-up II, but negatively mediated the effect on executive functioning with an ACME of -.07 SD. Additionally, teacher motivation also positively mediated the treatment effect on numeracy, socioemotional development and executive functioning with an ACME of .01 SD respectively at follow-up II.

Interestingly, teacher knowledge mediated students' numeracy skills, which aligns with some of the earlier literature on teacher knowledge as an important predictor of mathematics achievement (Lee, 2017; Garet et al., 2016; Thames, 2010). Implementation quality had positively mediated the treatment effect on classroom quality, and some of it trickled down to students' literacy skills. There is abundant research that shows the quality of teaching practices, and more specifically the implementation quality as explained earlier, is closely related to strong student learning and development across academic domains (Dobbie & Fryer, 2013; Hamre et al., 2014; Hamre, 2014; Mujis et al., 2014; Domitrovich & Greenberg, 2004). However, implementation quality has a negative mediation effect on executive functioning. This is rather strange but it appears that improvement in executive functioning requires implementation practices that may be

very different, or opposite of those required to improve literacy. Research shows that poor implementation quality may also yield negative effects on student outcomes (Clotfelter et al., 2007). Further, teacher motivation positively mediated the several student outcomes at follow-up II. This is particularly reassuring since many studies have shown how motivation is a critical determinant of student outcomes (Collie & Martin, 2017; Klassen et al., 2012; Martin, 2009). However, the results were fairly sensitive to the presence of unobserved omitted pretreatment confounders. These may include a combination of external family or household characteristics and parent-child interactions along with parent-teacher relationships that influenced the mediators and student outcomes (Kabay et al., 2017; Connor et al., 2005). Recent research from Malawi shows that strong parent-teacher partnership programs can be particularly influential in improving developmental outcomes amongst children (Ozler et al., 2018). Other research shows that teachers' own set of personal and professional risk factors also influence their professional well-being and student learning (Fatima & Wolf, 2020).

Furthermore, a small positive mediation effect on literacy and a negative mediation effect on executive functioning persisted when multiple mediators were examined together. Note that some of the omitted pretreatment confounders such as out-of-school factors impacting teacher and students' lives may also explain the negative mediation effect on executive functioning. However, since we are testing multiple hypotheses, there is a likelihood of finding an erroneous inference, just due to chance.

Finally, the mediation effects are extremely small on student outcomes, compared to those on classroom quality. This is not surprising since the original impact study by Wolf

et al. (2018a) and other studies (e.g. Egert et al., 2018) have also illustrated a similar pattern. As described earlier, improved implementation quality along with attention to omitted confounders such as family and home characteristics can result in pronounced mediation and overall treatment effect on student learning and development (Durlak & DuPre, 2008; Connor et al., 2005).

1.10.3. Limitations and Conclusions

The findings of this study provide a framework to understand the role of potential mechanisms that mediate the treatment effect of teacher training and professional development programs on classroom quality and student outcomes. However, there are a few limitations to keep in mind. First, the sample of the study is limited to six urban districts of the Greater Accra Region in Ghana, which limits the generalizability of the results. Second, the mediation effects on student outcomes are fairly sensitive to presence of unobserved omitted pretreatment confounders that limit the extent of inference. Third, the results are based on two snapshots across time and do not provide a continuous understanding of mediating mechanisms at play.

Nevertheless, this study has important implications for practitioners and policy makers with regards to the design of teacher trainings and professional development programs. Often times, professional development takes place in isolation of the actual implementation of teaching practices (Caspary, 2002). Results suggest that the implementation quality is a primary mediator of treatment effect on classroom quality and therefore, it is important to pivot professional development programs accordingly. These may include programs that incorporate active learning and modeling techniques

(Darling-Hammond et al., 2017) or provide ongoing feedback and mentorship to improve implementation quality and instructional practices (Guskey, 2002; Guskey & Yoon, 2009; Kraft, Blazar, & Hogam, 2018). However, the challenge remains to comprehensively explicate the mechanisms that mediate the impacts on student learning and development.

PART 2: WHAT TYPES OF TEACHER PROFILES INFLUENCE STUDENT LEARNING?

2.1. Introduction

The landmark Sustainable Development Goals have increasingly emphasized the role teachers play in facilitating positive student learning and development. However, with this renewed realization, researchers, practitioners and policy makers have become more interested in identifying rigorous analytical approaches and developing new valid measures to quantify teacher quality and associated student learning.

Within the field of education and economics, TVA approaches have been used widely to understand the relationship between teacher quality and student learning outcomes (Chetty et al., 2014; Kane & Staiger, 2008; Bau & Das, 2007; Rockoff, 2004). In other disciplines, specifically at the intersection of education and psychology, classroom observation tools have existed for over three decades (Gage & Needels, 1989). Some of these focus extensively on specific teacher behaviors, such as using frequency counts to evaluate the quantity of teaching (Cochran-Smith & Lytle, 1990; Smith, Waller, & Waller, 1982), while others have focused on teachers' time allocation (Brophy & Evertson, 1976; Fisher et al., 1980; Stallings, 1975), and classroom management behaviors (Coker, Medley, & Soar, 1980).

Over time, many more have been developed and validated as more sophisticated procedures have evolved to measure teacher and classroom quality. Classroom Assessment Scoring System (CLASS; Pianta et al., 2008) is a standardized measure of global classroom quality in three domains: emotional supports, classroom organization

and instructional supports. It can be used from prekindergarten through 12th grade. Similarly, Framework for Teaching (FFT; Danielson, 2013) is another classroom observation tool, in which the teaching activity is divided into 22 components under four domains: planning and preparation, professional responsibilities, classroom environment, and instruction. FFT has been widely used in numerous school districts across the United States. Protocol for Language Arts Teaching Observations (PLATO; Grossman et al., 2013) was developed to assess quality of teaching in English Language Arts specifically. It measures teaching practices in three domains: disciplinary demand, instructional scaffolding and classroom environment. The reliability coefficients for total scores from these three instruments range between .31 and .37 (Halpin & Kieffer, 2015; Kane & Staiger, 2012), and quadrupling the number of administrations (i.e. test-retest reliability) will result in reliabilities in the range of .60 - .70 (Ho & Kane, 2013).

More recently, Teacher Instructional Practices and Processes System (TIPPS; Seidman et al., 2017) and Teach (Molina et al., 2018) have been used to measure general constructs of teacher quality, specifically in low-resource contexts. TIPPS measures teaching practices in three domains: facilitating deeper learning, emotional support and behavior management and supporting student expression (as explained in Measures in Part I). Similarly, Teach measures quality of teaching practices in three domains as well: classroom culture, instruction and socioemotional skills. Average scores from these instruments have demonstrated high internal consistency, with reliability coefficients close to or greater than .70.

However, scores of teacher and classroom quality measured by these instruments are often presented as an aggregate at the construct or domain level. In some cases, they might even be presented at item level which yields an altogether different set of problems of reliability and validity. Nevertheless, either approach is in contrast to the notion of teacher competencies framework (Blömeke & Delaney, 2012), and other studies that have found that effective teaching is best understood as a function of multiple constructs that interactively influence classroom quality and student outcomes (Kraft et al., 2017; Conn, 2014). Therefore, to better understand teacher practices as measured by these instruments in a holistic manner, this study clusters teachers into subgroups (instead of items into constructs) across a set of professional well-being characteristics and instructional practices, and examines how each subgroup influences student learning. This classification allows us to identify the instructional needs of teachers within each subgroup and provide formative feedback with targeted interventions specifically designed for each subgroup. Clustering teachers instead of items facilitates more specific and actionable feedback to efficiently improve teaching quality and student learning.

2.2. Teacher profiles

Previous studies on classroom observation tools have typically found three overarching constructs of teaching practices (Kane & Staiger, 2012; Grossman et al., 2013; Lazarev & Newman, 2014), namely classroom environment, instructional quality and socio-emotional support. Most of these studies have relied on traditional psychometric approaches including factor analyses and item response theory (IRT) to assess dimensionality and item-level parameters of these measurement tools. However,

Halpin and Keiffer (2015) moved away from these approaches to adopt a model-based clustering strategy, commonly known as latent class analysis (LCA) to identify subgroups of teachers with similar teaching practices. The authors found four latent classes that corresponded to four distinct profiles of instructional practices. They concluded that neither profile fit into only one construct or domain of measurement and differentially predicted student learning. For example, certain profiles were more adept at facilitating student engagement and socioemotional development compared to other profiles. As noted by the authors, this allows "development of interventions that are both targeted at the needs of individual teachers and coordinated across multiple domains of practice" (Halpin & Keiffer, 2015). Another recent study used a similar person-oriented clustering strategy to understand work burnout and engagement among teachers (Salmela-Aro, Hietajarvi & Lonka, 2019). These studies suggest that certain groups of teachers are perhaps more likely to influence positive student learning outcomes than others, and therefore providing targeted interventions to help these groups of teachers improve in their lacking competencies can be particularly useful.

2.3. Targeted interventions

Many targeted interventions have been designed specifically for students with different educational backgrounds and learning levels over time. For example, Response to Intervention (RtI) is a multi-tier approach for early identification and provision of support to students with learning and behavioral needs (Burns & Gibbons, 2013; Jimerson et al., 2007; Fuchs & Fuchs, 2006; Bradley et al., 2005). Tier 1 includes high-quality classroom instruction, screening and group intervention, Tier 2 includes targeted

interventions for students not making progress and Tier 3 includes intensive interventions with comprehensive evaluations. It allows teachers to provide personalized support to group of students and within them, identify outliers that require more intensive support systems. The approach has been widely used across school systems in the United States. A similar educational approach, commonly known as Teaching at the Right Level or TaRL to help children develop basic reading and mathematics skills, especially in low- and middle-income countries (Banerjee et al., 2016). The approach works by dividing children into groups based on learning needs rather than age or gender, and dedicating time to basic skills rather than focusing on curriculum. Further, it entails regularly assessing student performance, instead of only during exam periods.

Similar to these strategies where instruction is tailored to the learning needs of the child, supports can be offered to teachers based on their teaching needs, rather than a "one size fits all" training or professional development programs. Some researchers have described these professional development activities as collaborative yet differentiated for individual teaching needs (Chambers et al., 2008; Portin et al, 2006). Other practitioners have described these as a paradigm shift away from a one-time professional development workshop to ongoing professional learning, which has been personalized based on the strengths and needs of the teachers, grounded in the tenets of adult learning theory, sustained and supported through implementation with coaching and follow-up, and consistently evaluated for its impact on student learning and adjusted when necessary (Moir, 2015; WGU, 2019). This aligns with doing professional development *with* the teachers, instead of *to* them (Ferlazzo, 2018).

2.4. Study

This study examines teacher profiles of instructional strengths and weaknesses. Building on earlier research on competency frameworks and targeted interventions, it specifically addresses the following research questions:

1. Based on constructs of professional well-being and instructional practices, what type of different teacher profiles exist?
2. How do these teacher profiles relate to student learning outcomes in the same year?
3. Are these profiles predicting differential treatment effect in follow-up I and follow-up II?

By identifying these profiles, a needs-based approach can be used to assess growth opportunities and provide targeted interventions for different subgroups of teachers. Moreover, future research, preferably experimental studies, will help better understand the efficacy of these targeted interventions.

2.5. Sample

To answer the first research question, the full sample of 444 teachers at baseline was utilized. Similarly, to answer the second research question, the full sample of 3435 students at baseline was used. However, to answer the third research question, the baseline clusters were examined for differential treatment effect on student learning outcomes at follow-up I and follow-up II. Therefore, the student sample comprised of 2358 students at follow-up I and 1883 students at follow-up II. Refer to Sample in Part I for additional details.

2.6. Measures

2.6.1. Profiles

Standardized measures of teachers' professional well-being and instructional practices (also referred to as measures of classroom quality) at baseline were used to classify teachers. Teachers' professional well-being was measured using teacher motivation, burnout and job satisfaction. Teachers' instructional practices were assessed using measures for facilitating deeper learning, emotional support and behavior management, and supporting student expression. Note teacher knowledge was not measured at baseline and therefore was not used in the analyses for this part of the study. All measures were comparably scaled, theoretically balanced and maintained factorial validity. Refer to Measures in Part 1 for additional details.

Correlational estimates show there is low correlation between different domains of professional well-being, ranging between .11 and .33, and low to moderate correlation between different domains of instructional practices, ranging between .12 and .45, at baseline, follow-up I and follow-up II. The low to moderate correlations show that each domain is adding unique value to distinctly classify teachers into different subgroups. Scores for most of these constructs (except for facilitating deeper learning) demonstrate high internal consistency, with Cronbach's alpha close to or greater than .70.

2.6.2. Criterion variables

Standardized measures of student outcomes were used to validate the teacher profiles. These included scores on students' numeracy skills, literacy skills,

socioemotional development and executive functioning. Refer to Measures in Part 1 for additional details.

2.7. Methods

Profile development. Cluster analysis was used to sort the 444 teachers into mutually exclusive groups. The technique takes into account the complexity of the data, categorizing the teachers based on properties of level, shape and dispersion, thereby ensuring distinctiveness, replicability and full coverage. Distinctiveness ensures that the teachers within each cluster are maximally similar to one another and dissimilar to those in alternative clusters. Replicability ensures that the clusters are not merely an artifact of chance, thus reducing the likelihood of a sampling error driving the cluster solution. Full coverage ensures that the resulting cluster solution is representative (Alterman et al., 1998).

Amongst numerous clustering algorithms, Ward's (1963) minimum variance procedure was best suited to the research objectives. Previously, Monte Carlo studies have compared competing clustering methods and have shown that when full coverage is required, Ward's method better recovers the known structure (Fisher, 1975; Mojena, 1977) and minimizes overlap (Bayue et al., 1980).

A three-stage clustering process was applied. First, raw scores for the six teacher attributes – motivation, burnout, job satisfaction, facilitating deeper learning, emotional support and behavior management, and supporting student expression were converted to standardized scores ($M = 0, SD = 1$), and the 444 teachers were randomly partitioned into four mutually exclusive blocks of 111. Ward's minimum variance procedure was

applied independently to the teachers comprising each block. For each block, the ideal number of clusters was determined through multiple criteria as used in earlier studies (e.g., Konold et al., 1999; Alterman et al., 1998). The criteria included a) atypical decrease in the overall between-cluster variance (R^2) and increase in the within-cluster variance (Ward, 1963) with no reverse trend at subsequent steps, b) ceiling of <1.0 for the ratio of within-cluster variances to variance for the full supply of cases within each block, and c) simultaneous elevation of the pseudo-*F* statistic over the pseudo-s^2statistic (Duda & Hart, 1973).

Clusters derived from the four independent first-stage analyses were pooled for second-stage clustering. Specifically, a similarity matrix was constructed using first-stage cluster mean-profiles, radial and dispersion statistics, and within-cluster profile frequency. The similarity matrix was subjected to the same Ward's procedure as before. This allowed first-stage clusters to provide independent replications of the final cluster solution. However, agglomerative clustering does not allow relocation of profiles retrospectively found to be misplaced. Therefore, third-stage clustering with divisive k means iteration was used to relocate any misplaced profiles. The selection criteria for second- and third-stage clustering was the same as for first-stage clustering with an additional set of more conservative stopping rules, including a) the average within-type homogeneity coefficient, $\overline{H}$ (Tryon & Bailey, 1970), must be $\geq .60$ (per McDermott et al., 1989); b) the average between-types similarity coefficient, $\overline{r_p}$ (Cattell, 1949), must be $< .40$ (also McDermott et. al, 1989); c) each final cluster should have a $\geq 50\%$ replication rate as verified by absorption of the first-stage cluster into the same second-

and third- stage cluster (as per Overall & Magee, 1992) and d) the solution must make psychological sense in terms of parsimonious coverage of the data and compatibility with teacher profile research (Alterman et al., 1998).

Profile explication. Internal and external variables were used to characterize and support the validity of the final profiles. Internal variables included constructs of classroom quality and professional well-being at baseline. External or criterion variables included student learning outcomes in numeracy, literacy, socioemotional development and executive functioning at baseline. Student learning outcomes were regressed on profile types, controlling for school (private/public ownership of the school), teacher (gender, age, experience) and student (gender, age) characteristics.

Additionally, how the profiles were moderated by the treatment to predict student learning at follow-up I and follow-up II was examined to identify whether certain profiles were more likely to benefit from the teacher training program compared to others, and may suggest different support interventions for different types of teacher profiles. Again, the same set of school, teacher and student controls was used.

2.8. Results

2.8.1. Profile structure

First-stage clustering produced 31 profile groups (an average of 8 per block). These were submitted to second-stage clustering based on a 31 X 31 similarity matrix and the solution was evaluated against the stated criteria. A six-cluster solution was found, which was submitted to a third-stage clustering with iterative partitioning for misplaced profile relocation.

Table 3 presents the average coefficient for within-profile homogeneity, between-profile similarity, and replication rate. The six-cluster solution yielded an exceedingly impressive tightness of fit for profiles overall, with the average $\bar{H}$ of .99 well above the a priori criterion of $\geq .60$. Similarly, there was reasonably high distinctiveness between respective profiles with the average $\bar{r_p}$ of .19 below the $< .40$ criterion. Moreover, the first four profiles were replicated perfectly (100%) across the four independent blocks, while the fifth profile was replicated 75% of the time and the sixth profile was replicated 50% of the time, satisfying the $\geq$ 50% criterion.

Table 3: Descriptive statistics of teacher profiles.

Type	Prevalence	Within-profile homogeneity (H)[a]	Between-profile similarity[b] (r_p)	% replicability across 4 independent blocks[c]
1	14.19	.99	.30	100
2	14.41	.99	.23	100
3	11.04	.99	.13	100
4	13.29	.99	.18	100
5	12.84	.99	.27	75
6	3.15	.99	.00	50
		$\bar{H} = .99$	$\bar{r_p} = .19$	

Note. $N = 444$.

[a]Within-profile homogeneity reveals the degree of profile similarity among the teachers comprising each profile. An H value of 1.0 would indicate that all teachers within a given profile are identical. H decreases as the variability of teachers within a given profile increases. An H value of 0.0 would indicate that the variability of teachers within a given profile equals the variability of teachers within the entire sample.

[b]Between-profile similarity indicates the degree of similarity between the mean attributes of a profile and the mean attributes of all other profiles. An r_p of 1.0 indicates that the mean attributes of a profile are identical to the mean attributes of all other profiles. As r_p decreases, the similarity between the mean attributes of a profile and all others decreases.

[c]Replicability of final profile is identified by examining whether the profile existed within each of the four first-stage cluster solutions. The percentage corresponds to the number of the first-stage solutions in which the final profile was found.

2.8.2. Profile membership

Table 4 presents a profile description with standardized mean scores for each of the six attributes within each profile. The profile description has been assigned based on the mean scores for attributes of professional well-being and instructional practices.

Table 4: Characterization of teacher profiles using internal variables.

Third Stage Clusters						
Profile	N	Variable	Standardized Mean	SD	Min	Max
Profile 1: Average professional well-being and classroom quality	63	Burnout	-0.53	0.54	-1.16	1.32
		Job Satisfaction	-0.53	0.75	-1.33	1.58
		Motivation	0.24	0.44	-1.14	0.58
		Emotional support and behavior management	0.33	0.81	-1.44	1.79
		Facilitating deeper learning	0.90	0.53	-0.08	1.87
		Supporting student expression	-0.25	0.54	-0.93	0.99
Profile 2: Good professional well-being but very poor classroom quality	64	Burnout	-0.15	0.83	-1.16	2.31
		Job Satisfaction	0.29	0.82	-1.33	2.90
		Motivation	0.12	0.59	-2.17	0.58
		Emotional support and behavior management	-1.29	0.50	-2.74	-0.15
		Facilitating deeper learning	-0.75	0.58	-1.55	0.89
		Supporting student expression	-0.28	0.64	-0.93	1.47
Profile 3: Good professional well-being and excellent classroom quality	49	Burnout	-0.29	0.49	-1.16	0.73
		Job Satisfaction	0.03	1.06	-1.33	3.17
		Motivation	0.10	0.61	-1.48	0.58
		Emotional support and behavior management	0.52	0.76	-1.12	1.79
		Facilitating deeper learning	1.00	0.59	-0.08	1.87
		Supporting student expression	1.61	0.64	0.51	2.91
Profile 4: Average professional well-being but poor classroom quality	59	Burnout	1.38	0.73	-0.46	3.11
		Job Satisfaction	0.76	0.88	-0.8	3.17
		Motivation	0.02	0.75	-2.86	0.58
		Emotional support and behavior management	0.38	0.79	-1.44	1.79
		Facilitating deeper learning	-0.19	0.81	-1.55	1.87
		Supporting student expression	-0.48	0.54	-0.93	0.99
Profile 5: Average professional well-being but very poor classroom quality	57	Burnout	-0.43	0.68	-1.16	1.22
		Job Satisfaction	-0.71	0.57	-1.33	0.78
		Motivation	0.22	0.54	-1.48	0.58
		Emotional support and behavior management	0.30	0.74	-1.12	1.47
		Facilitating deeper learning	-0.8	0.53	-1.55	-0.08
		Supporting student expression	-0.66	0.41	-0.93	0.51
Profile 6: Poor	14	Burnout	-0.36	0.63	-1.16	0.63
		Job Satisfaction	0.70	1.01	-1.33	2.64

professional well-being and average classroom quality	Motivation	-2.56	0.98	-4.23	-1.14
	Emotional support and behavior management	-0.30	0.60	-1.12	0.50
	Facilitating deeper learning	0.05	0.68	-1.55	0.89
	Supporting student expression	0.27	0.98	-0.93	1.71

Profile 1: Average professional well-being and classroom quality. This group of teachers, constituting slightly over 14% of the sample, indexed relatively high on motivation ($z = .24,\ SD = .44$) with low burnout ($z = -.53,\ SD = .54$). It was particularly adept at facilitating deeper learning ($z = .90,\ SD = .53$) and emotional support and behavior management ($z = .33,\ SD = .81$). However, this group had relatively low job satisfaction ($z = -.53,\ SD = .75$) and did not support student expression well ($z = -.25,\ SD = .54$). Overall, these teachers can be described as having average professional well-being and classroom quality.

Profile 2: Good professional well-being but very poor classroom quality. This group of teachers, again constituting over 14% of the sample, indexed high on professional well-being with relatively high motivation ($z = .12,\ SD = .59$) and job satisfaction ($z = .29,\ SD = .82$), and low burnout ($z = -.15,\ SD = .83$). Despite positive well-being, teachers exhibited poor instructional practices with extremely low scores on emotional support and behavior management ($z = -1.29,\ SD = .50$), facilitating deeper learning ($z = -.75,\ SD = .58$) and supporting student expression ($z = -.28,\ SD = .64$). Perhaps, these were just "lazy" teachers who are happy with their jobs because they do not need to work hard. Overall, these teachers can be described as having good professional well-being but still very poor classroom quality.

Profile 3: *Good professional well-being and excellent classroom quality.* This group of teachers, constituting 11% of the sample, had relatively high motivation ($z = .10,\ SD = .61$), average job satisfaction ($z = .03,\ SD = 1.06$) with low burnout ($z = -.29,\ SD = .49$). These teachers were particularly skilled at instructional practices with extremely high scores on supporting student expression ($z = 1.61,\ SD = .64$), facilitating deeper learning ($z = 1.00,\ SD = .59$), and emotional support and behavior management ($z = .52,\ SD = .76$). Overall, these teachers can be described as having good professional well-being and excellent classroom quality.

Profile 4: *Average professional well-being but poor classroom quality.* This group of teachers, constituting slightly over 13% of the sample, had average motivation ($z = .02,\ SD = .75$) and relatively high job satisfaction ($z = .76,\ SD = .88$), but extremely high burnout ($z = 1.38,\ SD = .73$). The group was relatively skilled at emotional support and behavior management ($z = .38,\ SD = .79$), but indexed low on supporting student expression ($z = -.48,\ SD = .54$) and facilitating deeper learning ($z = -.19,\ SD = .81$). Perhaps, these teachers burn out themselves providing emotional and behavior support that they are no longer able to facilitate student expression or deep learning. Overall, these teachers can be described as having average professional well-being but poor classroom quality.

Profile 5. *Average professional well-being but very poor classroom quality.* This group of teachers, constituting approximately 13% of the sample, had relatively high motivation ($z = .22,\ SD = .54$), low burnout ($z = -.43,\ SD = .68$) but poor job satisfaction ($z = -.71,\ SD = .57$). In terms of professional well-being, these teachers

appeared to be very similar to teachers in *Profile 1*. In terms of instructional practices, this group was relatively skilled at emotional support and behavior management ($z = .30,\ SD = .74$), but indexed very low on facilitating deeper learning ($z = -.80,\ SD = .53$) and supporting student expression ($z = -.66,\ SD = .41$). This is in contrast to teachers in *Profile* 1, who were also very skilled at facilitating deeper learning. Overall, these teachers can be described as having average professional well-being but very poor classroom quality.

Profile 6. *Poor professional well-being and average classroom quality* This group of teachers, constituting approximately a mere 3% of the sample, had relatively high job satisfaction ($z = .70,\ SD = 1.01$) and low burnout ($z = -.36,\ SD = .63$), but extremely low motivation up to two and a half standard deviation below average ($z = -2.56,\ SD = .98$). Teachers were able to support student expression ($z = .27,\ SD = .98$) and moderately facilitate deeper learning ($z = .05,\ SD = .68$) but indexed low on emotional support and behavior management ($z = -.30,\ SD = .60$). Overall, these teachers can be described as having poor professional well-being and average classroom quality.

2.8.3. Relationship with student learning

To establish the validity of the six-cluster solution, the relationship of the clusters with student learning outcomes at baseline was examined. Table 5 presents the regression estimates of student learning outcomes predicted by profile types, controlling for a set of school, teacher and student characteristics.

All estimates are relative to the *Profile 2* as the base category since teachers in this profile performed most poorly across all instructional practices. Teachers in *Profile 2* were described to have good professional well-being but very poor classroom quality.

Teachers in *Profile 1* (average professional well-being and classroom quality) were more likely to have students with higher numeracy and literacy skills by .22-.23 SD ($p < .1$) and higher socioemotional development by .18 SD ($p < .1$). Teachers in *Profile 3* (good professional well-being and excellent classroom quality) were more likely to have students with higher numeracy skills by .32 SD ($p < .001$) and literacy skills by .24 SD ($p < .1$). Teachers in all other profiles were not likely to perform better or worse than teachers in *Profile 2*.

Table 5. Relationship of teacher profiles with student learning outcomes at baseline (Base Category—Profile 2)

	Numeracy	Literacy	Socioemotional Development	Executive Functioning
Profile 1: Average professional well-being and classroom quality	.22*	.23*	.18*	.13
	(.11)	(.12)	(.10)	(.10)
Profile 3: Good professional well-being and excellent classroom quality	.32***	.24*	.12	.11
	(.12)	(.13)	(.10)	(.10)
Profile 4: Average professional well-being but poor classroom quality	.16	.07	.08	.03
	(.12)	(.13)	(.11)	(.10)
Profile 5: Average professional well-being but very poor classroom quality	-.12	-.09	-.12	-.11
	(.13)	(.13)	(.11)	(.11)
Profile 6: Poor professional well-being and average classroom quality	.22	.12	.10	.24
	(.21)	(.17)	(.14)	(.18)
School Controls	□	□	□	□

Teacher Controls	□	□	□	□
Student Controls	□	□	□	□
Observations	2,390	2,390	2,390	2,390
R-squared	.05	.07	.02	.02

Note. All coefficients estimates are relative to the base category – Profile 2. Teachers in Profile 2 have good professional well-being but very poor classroom quality. Robust standard errors in parentheses. School controls include private/public ownership of schools. Teacher controls include teachers' age, gender and years of experience. Student controls include students' age and gender.
*** p<0.01, ** p<0.05, * p<0.1

The analyses were repeated to understand estimates relative to Profiles 4 and 5 as well since teachers in these profiles also performed poorly across a subset of well-being traits and instructional practices (Tables 6 and 7). Teachers in *Profile 4* were described to have average professional well-being but poor classroom quality, while teachers in *Profile 5* were described to have average professional well-being but very poor classroom quality. Relative to teachers in *Profile 4*, teachers in *Profile 2* performed worse in numeracy by .21 SD ($p < .1$) while teachers in *Profile 5* performed worse in numeracy by .31 SD ($p < .05$) and socioemotional development by .22 SD ($p < .1$).

Table 6. Relationship of teacher profiles with student learning outcomes at baseline (Base Category—Profile 4)

	Numeracy	Literacy	Socioemotional Development	Executive Functioning
Profile 1: Average professional well-being and classroom quality	.04	.11	.08	0.11
	(.12)	(.13)	(.119)	(0.11)
Profile 2: Good professional well-being but very poor classroom quality	-.21*	-.15	-.12	-0.02
	(.12)	(.13)	(.111)	(0.10)
Profile 3: Good professional well-being and excellent classroom quality	.14	.13	.03	0.08
	(.13)	(.13)	(.11)	(0.12)
Profile 5: Average professional well-being	-.31**	-.20	-.22*	-0.14

but very poor classroom quality				
	(.13)	(.13)	(.12)	(0.13)
Profile 6: Poor professional well-being and average classroom quality	.04	.01	.01	0.21
	(.22)	(.18)	(.15)	(0.20)
School Controls	□	□	□	□
Teacher Controls	□	□	□	□
Student Controls	□	□	□	□
Observations	2,390	2,390	2,390	2,390
R-squared	.05	.07	.02	0.02

Note. All coefficients estimates are relative to the base category – Profile 4. Teachers in Profile 4 have average professional well-being but poor classroom quality. Robust standard errors in parentheses. School controls include private/public ownership of schools. Teacher controls include teachers' age, gender and years of experience. Student controls include students' age and gender.

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

Relative to teachers in *Profile 5*, teachers in *Profile 1* performed better in numeracy by .31 SD ($p < .05$), literacy by .26 SD ($p < .05$), socioemotional development by .27 SD ($p < .05$) and executive functioning by .24 SD ($p < .05$). Similarly, teachers in *Profile 3* performed better in numeracy by .41 SD ($p < 0.01$), literacy by .28 SD ($p < 0.05$), socioemotional development by .22 SD ($p < 0.05$), and executive functioning by .21 SD ($p < 0.05$). Additionally, teachers in *Profile 4* performed better in numeracy by .24 SD ($p < 0.1$) and teachers in *Profile 6* performed better in executive functioning by .34 SD ($p < 0.1$). This suggests that teachers in *Profile 5* performed the worst, followed by teachers in *Profile 2* and *Profile 4*, respectively.

Table 7. Relationship of teacher profiles with student learning outcomes at baseline (Base Category—Profile 5)

	Numeracy	Literacy	Socioemotional Development	Executive Functioning

Profile 1: Average professional well-being and classroom quality	.31**	.26**	.27**	.24**
	(.13)	(.13)	(.11)	(.12)
Profile 2: Good professional well-being but very poor classroom quality	.07	-.00	.08	.11
	(.13)	(.13)	(.11)	(.11)
Profile 3: Good professional well-being and excellent classroom quality	.41***	.28**	.22**	.21*
	(.13)	(.13)	(.11)	(.12)
Profile 4: Average professional well-being but poor classroom quality	.24*	.11	.17	.13
	(.13)	(.13)	(.12)	(.13)
Profile 6: Poor professional well-being and average classroom quality	.31	.16	.19	.34*
	(.22)	(.18)	(.15)	(.20)
School Controls	□	□	□	□
Teacher Controls	□	□	□	□
Student Controls	□	□	□	□
Observations	2,390	2,390	2,390	2,390
R-squared	0.05	0.07	0.02	0.02

Note. All coefficients estimates are relative to the base category – Profile 5. Teachers in Profile 5 have average professional well-being but very poor classroom quality. Robust standard errors in parentheses. School controls include private/public ownership of schools. Teacher controls include teachers' age, gender and years of experience. Student controls include students' age and gender.
*** p<0.01, ** p<0.05, * p<0.1

2.8.4. Differential treatment effects

The profiles did not exhibit any significant differential treatment effects on student learning outcomes. This means that no profile was particularly more receptive to the teacher training/coaching intervention. This suggests that the intervention was perhaps geared to the full group, and segmenting the intervention to focus on specific instructional practices can ensure all teachers benefit equitably to index highly on all domains of professional well-being and instructional practices.

2.9. Discussion

This study examined different subpopulations of teachers based on their instructional practices and professional well-being. Teachers were specifically classified into categories to understand their teaching needs to provide formative feedback and targeted interventions for support. Six teacher profiles were found: 1) teachers with average professional well-being (except job satisfaction) and average classroom quality (except supporting student expression) 2) teachers with good professional well-being but poor classroom quality 3) teachers with good professional well-being and excellent classroom quality 4) teachers with average professional well-being (except burnout) but poor classroom quality (except emotional support and behavior management) 5) teachers with average professional well-being (except job satisfaction) and very poor classroom quality (except emotional support and behavior management), and 6) teachers with poor professional well-being (due to extremely low motivation) but average classroom quality (except emotional support and behavior management).

Except for teachers in Profile 3, teachers in every profile were lacking in either a measure of professional well-being, classroom quality or both. Estimates for correlation with student learning show that teachers in Profile 5 (average professional well-being but very poor classroom quality) performed the worst, followed by teachers in Profile 2 (good professional well-being but very poor classroom quality), and Profile 4 (average professional well-being but poor classroom quality). Estimates for correlation with student learning show that compared to teachers in Profile 5, teachers in Profiles 1 and 3 were more likely to have students with higher numeracy and literacy skills,

socioemotional development and executive functioning. Teachers in Profile 4 were more likely to have students with higher numeracy skills, and teachers in Profile 6 were more likely to have students with higher executive functioning. Teachers in Profile 2 were not likely to have any better student outcomes. This helps identify groups of teachers that need strong support and targeted interventions to positively influence student learning. In this case, teachers in Profiles 5 and 2 indexed extremely low on instructional practices across classroom quality. Therefore, providing teachers within these groups with more personalized interventions to help them improve their teaching practices (i.e. facilitating deeper learning, emotional behavior and management and supporting student expression) could be helpful. Further, teachers in Profile 5 also had very poor job satisfaction so facilitating complementary interventions to specifically target this particular dimension of professional well-being could be helpful. Research shows that presence of strong leadership, provision of more ownership in decision-making processes, and facilitation of structural supports are linked to improved job satisfaction amongst teachers (Sharma & Jyoti, 2006; Edinger & Edinger, 2018; Crisci et al., 2019). Overall, the results help identify needs within a group of teachers to facilitate the development of formative feedback and personalized interventions to help teachers improve in the lacking competency. This aligns with earlier research that recommends professional development activities to be collaborative yet differentiated to meet the individual needs of the teachers (Chambers et al., 2008; Portin et al, 2006).

Further, additional analyses showed that the teacher profiles did not predict any differential treatment effects. This implies that the teacher training/coaching program was

perhaps geared to a general group of teachers and the positive treatment effect was averaged across all teachers. While comprehensive training programs are effective (Simonsen et al., 2008; Simonsen et al., 2014; Simonsen, Myers & DeLuca, 2010), they do not necessarily ensure that every teacher maximally benefits from the program and in some cases, they are also resource-intensive. On the contrary, targeted professional development based on a needs-identification strategy can help achieve farthermost benefits for every teacher in a more efficient manner (Darling-Hammond et al., 2017; Moir, 2015). Further, once every teacher is able to teach at their highest potential, the effects will likely be more pronounced for student learning and development.

2.9.1. Limitations and Conclusions

While empirical results have been estimated, the overarching goal of the paper is to facilitate provision of effective feedback and support systems for teachers. There are a few caveats to consider. First, the approach must be replicated across samples in different contexts to examine if similar profiles of teachers emerge and if targeted interventions can be standardized by profile type for scalability. Second, the profiles have been characterized on pure constructs of professional well-being and instructional practices. There are other within-teacher characteristics such as personality traits, educational background and demographic characteristics that may also play a role in determining which profile teachers are eventually placed in, and how well they respond to the targeted interventions. Finally, there may also be some measurement error with weak factorial integrity for one of the constructs of professional well-being – motivation and instructional practices – facilitating deeper learning, respectively.

Overall, the findings suggest that teachers can be classified into subgroups based on a set of professional well-being traits and instructional practices. This facilitates a needs-based approach to identify growth opportunities for each subgroup and efficiently provide support systems and mentorship to address the specific pain points. This will ensure teachers maximally benefit to implement high quality teaching across all domains of instructional practices. This would also likely lead to more marked positive impacts on student learning. However, future research (preferably experimental studies) would be required to better understand the overall efficacy of these targeted interventions on teacher quality and student learning.

CONCLUDING REMARKS

In summary, the first part of the study focused on understanding the influential pathways that mediate positive impacts of teacher professional development programs. It found that the implementation quality positively mediated the treatment effect on classroom quality, but not so much on student outcomes. The second part of the study dived deep into understanding how to target interventions to ensure all teachers maximally benefit from them. It found that clustering teachers into different subgroups helps identify their teaching needs across different teacher competencies which facilitates provision of effective formative feedback and targeted support. Consequently, improved teacher performance across all domains of instructional practices would likely lead to more pronounced positive impacts on student learning. Future research must rigorously examine the effectiveness of these targeted interventions for conclusive evidence.

APPENDIX A: DESCRIPTIVE STATISTICS OF TEACHER- AND STUDENT-LEVEL MEASURES

Table A1: Descriptive statistics of teacher-level measures of knowledge, well-being and classroom quality for the raw dataset.

Variable	*N*	Mean	Std. Dev.	Min	Max
Baseline					
Motivation	444	4.66	0.58	1.00	5.00
Burnout	443	2.06	0.92	1.00	7.00
Job Satisfaction	443	1.84	0.63	1.00	3.83
Facilitating deeper learning	315	2.03	0.69	1.00	3.33
Emotional support and behavior management	315	2.77	0.45	1.57	3.57
Supporting student expression	315	1.48	0.52	1.00	3.00
Implementation checklist	345	2.23	1.85	0.00	8.00
Follow-up I					
Developmentally appropriate practice	348	4.75	0.30	2.83	5.00
Supporting social and emotional development	348	4.64	0.39	2.20	5.00
Family sensitive caregiving	348	4.14	0.62	2.00	5.00
Motivation	348	4.71	0.44	1.80	5.00
Burnout	348	2.01	0.89	1.00	5.82
Job Satisfaction	348	3.09	0.68	1.00	4.00
Facilitating deeper learning	345	2.31	0.58	1.00	3.33
Emotional support and behavior management	345	3.07	0.37	1.71	3.71
Supporting student expression	345	1.66	0.55	1.00	3.00
Implementation checklist	345	4.32	1.89	0.00	11.00
Follow-up II					
Developmentally appropriate practice	309	4.73	0.27	3.83	5.00
Supporting social and emotional development	309	4.63	0.35	3.40	5.00
Family sensitive caregiving	309	4.17	0.60	2.20	5.00
Motivation	309	4.63	0.47	2.20	5.00
Burnout	309	2.09	0.91	1.00	5.82
Job Satisfaction	309	3.04	0.68	1.00	4.00
Facilitating deeper learning	295	2.37	0.74	1.00	4.00
Emotional support and behavior management	295	2.93	0.37	1.86	3.71
Supporting student expression	295	1.86	0.64	1.00	3.50
Implementation checklist	295	3.46	1.59	0.00	8.00

Table A2: Descriptive statistics of teacher-level measures of knowledge, well-being and classroom quality for the imputed dataset at follow-up I.

Variable	*N*	Mean	Std. Dev.	Min	Max
Baseline					
Motivation	345	4.64	0.60	1.00	5.00
Burnout	345	2.04	0.92	1.00	7.00
Job Satisfaction	345	1.81	0.61	1.00	3.83
Facilitating deeper learning	345	1.86	0.68	1.00	3.33
Emotional support and behavior management	345	2.80	0.37	1.57	3.57
Supporting student expression	345	1.49	0.45	1.00	3.00
Implementation checklist	345	2.23	1.85	0.00	8.00
Follow-up I					
Developmentally appropriate practice	345	4.75	0.30	2.83	5.00
Supporting social and emotional development	345	4.64	0.39	2.20	5.00
Family sensitive caregiving	345	4.14	0.62	2.00	5.00
Motivation	345	4.71	0.44	1.80	5.00
Burnout	345	2.01	0.89	1.00	5.82
Job Satisfaction	345	3.08	0.68	1.00	4.00
Facilitating deeper learning	345	2.31	0.58	1.00	3.33
Emotional support and behavior management	345	3.07	0.37	1.71	3.71
Supporting student expression	345	1.66	0.55	1.00	3.00
Implementation checklist	345	4.32	1.89	0.00	11.00

Table A3: Descriptive statistics of teacher-level measures of knowledge, well-being and classroom quality for the imputed dataset at follow-up II.

Variable	*N*	Mean	Std. Dev.	Min	Max
Baseline					
Motivation	295	4.61	0.62	1.00	5.00
Burnout	295	2.09	0.92	1.00	7.00
Job Satisfaction	295	1.81	0.59	1.00	3.67
Facilitating deeper learning	295	1.88	0.67	1.00	3.33
Emotional support and behavior management	295	2.83	0.36	1.86	3.57
Supporting student expression	295	1.52	0.45	1.00	3.00
Implementation checklist	295	2.28	1.87	0.00	8.00
Follow-up I					
Developmentally appropriate practice	295	4.76	0.29	3.50	5.00
Supporting social and emotional development	295	4.64	0.40	2.20	5.00
Family sensitive caregiving	295	4.17	0.61	2.00	5.00
Motivation	295	4.70	0.45	1.80	5.00
Burnout	295	2.01	0.90	1.00	5.82
Job Satisfaction	295	3.08	0.68	1.00	4.00
Facilitating deeper learning	295	2.32	0.58	1.00	3.33
Emotional support and behavior management	295	3.07	0.37	1.71	3.57
Supporting student expression	295	1.66	0.54	1.00	3.00
Implementation checklist	295	4.35	1.89	0.00	10.00
Follow-up II					
Developmentally appropriate practice	295	4.73	0.28	3.83	5.00
Supporting social and emotional development	295	4.63	0.35	3.40	5.00
Family sensitive caregiving	295	4.18	0.59	2.20	5.00
Motivation	295	4.64	0.45	2.60	5.00
Burnout	295	2.09	0.92	1.00	5.82
Job Satisfaction	295	3.04	0.68	1.00	4.00
Facilitating deeper learning	295	2.37	0.74	1.00	4.00
Emotional support and behavior management	295	2.93	0.37	1.86	3.71
Supporting student expression	295	1.86	0.64	1.00	3.50
Implementation checklist	295	3.46	1.59	0.00	8.00

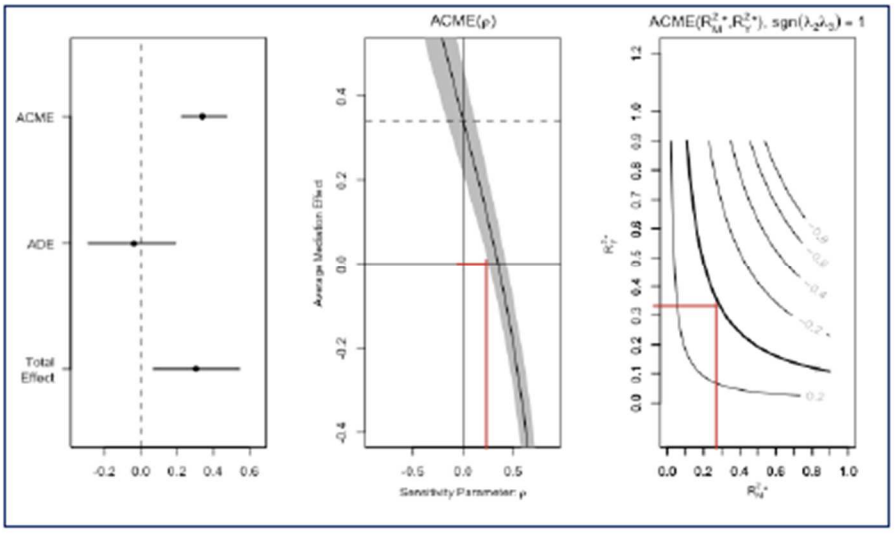
Table A4: Descriptive statistics of student-level measures of learning outcomes

Variable	*N*	Mean	Std. Dev.	Min	Max
Baseline					
Numeracy	3,435	0.44	0.19	0.00	0.99
Literacy	3,435	0.46	0.22	0.00	0.98
Social Emotional Development	3,435	0.41	0.2	0.00	0.97
Executive Functioning	3,435	0.49	0.21	0.00	0.89
Follow-up I					
Numeracy	2,358	0.58	0.19	0.00	1.00
Literacy	2,358	0.62	0.20	0.00	1.00
Social Emotional Development	2,358	0.55	0.19	0.00	0.97
Executive Functioning	2,358	0.59	0.18	0.00	0.89
Follow-up II					
Numeracy	1,883	0.68	0.16	0.08	1.00
Literacy	1,883	0.72	0.18	0.02	1.00
Social Emotional Development	1,883	0.59	0.17	0.07	0.97
Executive Functioning	1,883	0.64	0.16	0.00	0.89

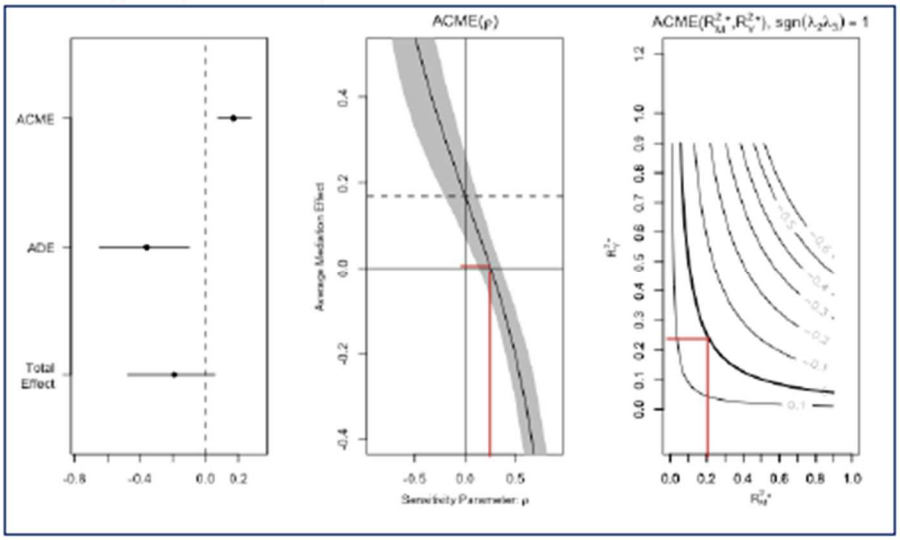
APPENDIX B: SENSITIVITY ANALYSES FOR MEDIATION EFFECTS ON CLASSROOM QUALITY

Figure B1: ACME, ADE and total effect for simple mediation by implementation quality of teaching practices on classroom quality.

Follow-up I (June 2016):



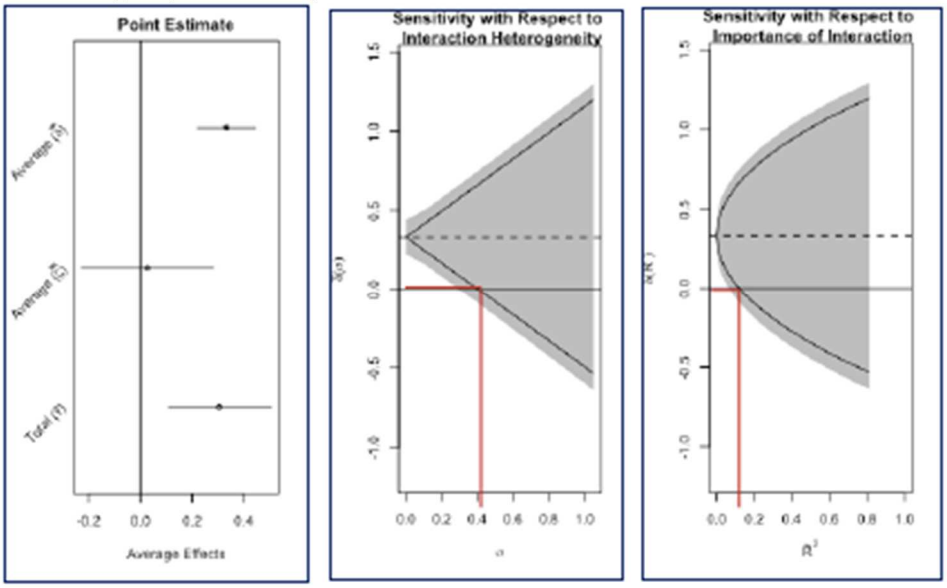
Follow-up II (June 2017):



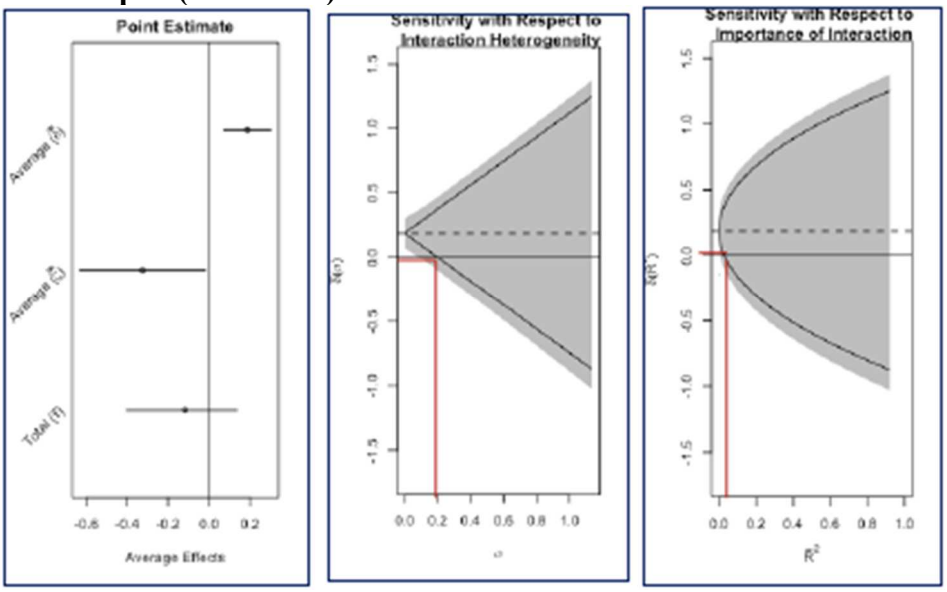
Note. The first figure from the left shows the estimate of ACME under the sequential ignorability assumption with the 95% bootstrap confidence interval. The middle figure shows the estimated ACME as function of the sensitivity parameter, ρ , which represents the correlation between the error terms in the mediator and the outcome models. The thick lines and gray bands represent the point estimates of the ACME and their 95% confidence intervals, respectively. The last figure shows the same sensitivity analyses, with the ACME plotted against (R_Y^2, R_M^2), the proportions of the total variance in the outcome and mediator variables, respectively, in the presence of unobserved pretreatment confounders. Overall, the causal mediation effects are positive and fairly robust to unobserved pretreatment mediator–outcome confounding to varying degrees.

Figure B2: ACME, ADE and total effect for multiple mediation by knowledge, implementation quality of teaching practices and professional well-being on classroom quality.

Follow-up I (June 2016):



Follow-up II (June 2017):



Note. The first figure from the left shows the estimate of the ACME under the sequential ignorability and homogeneous interaction assumptions with the 95% bootstrap confidence interval. The middle figure shows the sharp bounds on the ACME as a function of the sensitivity parameter, σ, which is the SD of the varying coefficient on the treatment–mediator interaction term and represents the degree of heterogeneity in the interaction. The last figure shows the ACME with respect to R^2, the proportion of the total variance of the outcome variable that would be explained by the treatment–mediator interaction term. Overall, the ACME is smaller in magnitude and weaker in significance compared to the simple mediation of implementation quality of teaching practices on classroom quality.

APPENDIX C: MEDIATION EFFECTS ON STUDENT OUTCOMES

Table C1: Estimates for simple mediation on student outcomes—numeracy skills.

	Follow-up I (June 2016)					Follow-up II (June 2017)				
	Estimate	95% CI Lower	95% CI Upper	p-value		Estimate	95% CI Lower	95% CI Upper	p-value	
Knowledge										
ACME	0.01	0.00	0.02	0.10	.	0.01	0.00	0.01	0.22	
ADE	0.03	-0.03	0.10	0.32		-0.03	-0.11	0.04	0.40	
Total Effect	0.04	-0.02	0.11	0.19		-0.03	-0.10	0.04	0.47	
Prop. Mediated	0.20	-1.89	3.39	0.28		-0.19	-1.85	1.47	0.58	
Implementation										
ACME	0.00	-0.03	0.03	0.99		0.02	0.00	0.04	0.08	.
ADE	0.04	-0.03	0.11	0.24		-0.05	-0.12	0.04	0.26	
Total Effect	0.04	-0.02	0.11	0.22		-0.03	-0.10	0.05	0.49	
Prop. Mediated	-0.02	-2.33	3.14	0.99		-0.65	-7.24	7.48	0.52	
Motivation										
ACME	0.00	0.00	0.01	0.48		0.00	0.00	0.01	0.03	*
ADE	0.04	-0.02	0.10	0.23		-0.03	-0.11	0.03	0.40	
Total Effect	0.04	-0.02	0.10	0.22		-0.03	-0.10	0.04	0.47	
Prop. Mediated	0.03	-0.39	0.35	0.58		-0.18	-1.73	1.74	0.49	
Job Satisfaction										
ACME	0.00	0.00	0.00	0.97		0.00	0.00	0.01	0.63	
ADE	0.04	-0.03	0.10	0.22		-0.03	-0.10	0.05	0.44	
Total Effect	0.04	-0.02	0.10	0.22		-0.03	-0.10	0.05	0.46	
Prop. Mediated	0.00	-0.08	0.13	0.95		-0.04	-0.76	0.51	0.87	
Burnout										
ACME	0.01	0.00	0.02	0.09	.	0.00	-0.01	0.00	0.36	
ADE	0.03	-0.03	0.10	0.32		-0.03	-0.10	0.05	0.50	
Total Effect	0.04	-0.03	0.10	0.22		-0.03	-0.10	0.05	0.48	
Prop. Mediated	0.20	-2.12	1.91	0.28		0.08	-0.75	0.69	0.67	

Note. Estimates under the sequential ignorability assumption.

Table C2: Estimates for simple mediation on student outcomes—literacy skills.

	Follow-up I (June 2016)				Follow-up II (June 2017)			
	Estimate	95% CI Lower	95% CI Upper	p-value	Estimate	95% CI Lower	95% CI Upper	p-value
Knowledge								
ACME	0.01	0.00	0.02	0.12	0.00	-0.01	0.00	0.47
ADE	0.02	-0.04	0.08	0.52	0.00	-0.07	0.07	0.97
Total Effect	0.03	-0.03	0.09	0.33	0.00	-0.08	0.07	0.97
Prop. Mediated	0.30	-3.07	3.15	0.42	1.44	-1.31	1.50	0.99
Implementation								
ACME	0.03	0.00	0.06	0.06 .	0.02	0.00	0.04	0.09 .
ADE	0.00	-0.07	0.07	0.96	-0.02	-0.09	0.05	0.62
Total Effect	0.03	-0.04	0.10	0.36	0.00	-0.08	0.07	0.97
Prop. Mediated	0.99	-8.69	9.49	0.40	-8.40	-8.90	8.01	0.99
Motivation								
ACME	0.00	-0.01	0.00	0.30	0.00	0.00	0.01	0.42
ADE	0.03	-0.03	0.09	0.33	0.00	-0.07	0.06	0.97
Total Effect	0.03	-0.04	0.09	0.36	0.00	-0.07	0.06	0.99
Prop. Mediated	-0.06	-0.59	0.65	0.55	-0.83	-0.66	0.65	0.99
Job Satisfaction								
ACME	0.00	0.00	0.00	0.60	0.00	0.00	0.00	0.80
ADE	0.03	-0.03	0.09	0.35	0.00	-0.08	0.06	0.92
Total Effect	0.03	-0.03	0.09	0.37	0.00	-0.07	0.06	0.94
Prop. Mediated	-0.02	-0.30	0.18	0.73	-0.22	-0.47	0.41	0.98
Burnout								
ACME	0.01	0.00	0.02	0.27	0.00	-0.01	0.00	0.70
ADE	0.02	-0.04	0.09	0.49	0.00	-0.07	0.07	1.00
Total Effect	0.03	-0.03	0.09	0.36	0.00	-0.07	0.07	0.97
Prop. Mediated	0.20	-1.91	2.73	0.53	0.41	-1.41	1.08	0.97

Note. Estimates under the sequential ignorability assumption.

Table C3: Estimates for simple mediation on student outcomes—socioemotional development.

	Follow-up I (June 2016)					Follow-up II (June 2017)				
	Estimate	95% CI Lower	95% CI Upper	p-value		Estimate	95% CI Lower	95% CI Upper	p-value	
Knowledge										
ACME	0.00	-0.02	0.01	0.66		0.01	0.00	0.02	0.22	
ADE	0.23	0.14	0.31	0.00	***	-0.03	-0.12	0.07	0.54	
Total Effect	0.22	0.14	0.30	0.00	***	-0.02	-0.11	0.07	0.63	
Prop. Mediated	-0.01	-0.07	0.06	0.66		-0.29	-2.75	1.66	0.70	
Implementation										
ACME	-0.01	-0.05	0.03	0.63		0.00	-0.02	0.03	0.76	
ADE	0.23	0.14	0.32	0.00	***	-0.03	-0.11	0.07	0.60	
Total Effect	0.22	0.14	0.30	0.00	***	-0.02	-0.11	0.07	0.65	
Prop. Mediated	-0.04	-0.23	0.12	0.63		-0.18	-2.79	3.95	0.90	
Motivation										
ACME	0.00	-0.01	0.01	0.95		0.01	0.00	0.02	0.02	*
ADE	0.22	0.14	0.30	0.00	***	-0.03	-0.12	0.06	0.52	
Total Effect	0.22	0.14	0.30	0.00	***	-0.02	-0.12	0.07	0.62	
Prop. Mediated	0.00	-0.03	0.03	0.95		-0.41	-5.02	2.55	0.64	
Job Satisfaction										
ACME	0.00	0.00	0.00	0.76		0.00	0.00	0.00	1.00	
ADE	0.22	0.14	0.30	0.00	***	-0.02	-0.11	0.07	0.66	
Total Effect	0.22	0.14	0.30	0.00	***	-0.02	-0.11	0.07	0.65	
Prop. Mediated	0.00	-0.02	0.01	0.76		0.00	-0.24	0.25	0.98	
Burnout										
ACME	0.01	0.00	0.02	0.16		0.00	-0.01	0.00	0.44	
ADE	0.21	0.13	0.30	0.00	***	-0.02	-0.11	0.08	0.74	
Total Effect	0.22	0.15	0.30	0.00	***	-0.02	-0.11	0.08	0.72	
Prop. Mediated	0.04	-0.01	0.11	0.16		0.14	-0.89	0.81	0.84	

Note. Estimates under the sequential ignorability assumption.

Table C4: Estimates for simple mediation on student outcomes—executive functioning.

	Follow-up I (June 2016)					Follow-up II (June 2017)				
	Estimate	95% CI Lower	95% CI Upper	p-value		Estimate	95% CI Lower	95% CI Upper	p-value	
Knowledge										
ACME	0.01	-0.01	0.02	0.46		0.00	-0.01	0.00	0.34	
ADE	0.10	0.02	0.18	0.01	*	-0.08	-0.17	0.01	0.09	.
Total Effect	0.11	0.03	0.19	0.01	**	-0.08	-0.17	0.01	0.08	.
Prop. Mediated	0.05	-0.12	0.29	0.46		0.05	-0.13	0.50	0.38	
Implementation										
ACME	-0.07	-0.11	-0.04	0.00	***	0.01	-0.01	0.04	0.24	
ADE	0.18	0.09	0.27	0.00	***	-0.10	-0.19	0.00	0.05	*
Total Effect	0.11	0.02	0.19	0.01	**	-0.08	-0.17	0.01	0.09	.
Prop. Mediated	-0.67	-3.06	-0.29	0.01	**	-0.18	-1.81	1.50	0.32	
Motivation										
ACME	0.00	0.00	0.01	0.06	.	0.00	0.00	0.01	0.44	
ADE	0.10	0.02	0.19	0.01	*	-0.08	-0.17	0.01	0.10	
Total Effect	0.11	0.03	0.19	0.01	**	-0.08	-0.17	0.02	0.11	
Prop. Mediated	0.04	0.00	0.20	0.06	.	-0.03	-0.28	0.18	0.52	
Job Satisfaction										
ACME	0.00	0.00	0.00	0.98		0.00	0.00	0.00	0.73	
ADE	0.11	0.03	0.19	0.00	***	-0.08	-0.17	0.01	0.08	.
Total Effect	0.11	0.03	0.19	0.00	***	-0.08	-0.17	0.01	0.09	.
Prop. Mediated	0.00	-0.03	0.02	0.98		-0.01	-0.15	0.05	0.73	
Burnout										
ACME	0.02	0.00	0.03	0.01	*	0.00	-0.01	0.00	0.16	
ADE	0.09	0.01	0.18	0.01	*	-0.08	-0.17	0.01	0.09	.
Total Effect	0.11	0.03	0.19	0.01	**	-0.08	-0.18	0.01	0.07	.
Prop. Mediated	0.14	0.02	0.56	0.02	*	0.06	-0.13	0.37	0.21	

Note. Estimates under the sequential ignorability assumption.

Table C5: Estimates for multiple mediation on student outcomes –numeracy skills.

	Follow-up I (June 2016)			Follow-up II (June 2017)		
	Estimate	95% CI Lower	95% CI Upper	Estimate	95% CI Lower	95% CI Upper
ACME (treated)	-0.01	-0.04	0.02	0.01	-0.01	0.04
ACME (control)	0.03	-0.04	0.09	0.03	-0.01	0.07
ACME (average)	0.00	-0.03	0.03	0.02	0.00	0.04
ADE (treated)	0.01	-0.08	0.11	-0.06	-0.14	0.02
ADE (control)	0.05	-0.03	0.12	-0.04	-0.12	0.04
ADE (average)	0.02	-0.06	0.10	-0.06	-0.13	0.02
Total Effect	0.04	-0.03	0.11	-0.03	-0.10	0.05

Note. Estimates under the sequential ignorability and homogeneous interaction assumptions.

Table C6: Estimates for multiple mediation on student outcomes –literacy skills.

	Follow-up I (June 2016)			Follow-up II (June 2017)		
	Estimate	95% CI Lower	95% CI Upper	Estimate	95% CI Lower	95% CI Upper
ACME (treated)	0.02	-0.02	0.05	0.00	-0.03	0.02
ACME (control)	0.10	0.03	0.16	0.06	0.02	0.10
ACME (average)	**0.04**	**0.01**	**0.07**	0.01	-0.01	0.04
ADE (treated)	-0.07	-0.16	0.02	-0.06	-0.14	0.02
ADE (control)	0.01	-0.06	0.08	0.00	-0.08	0.08
ADE (average)	-0.05	-0.12	0.03	-0.04	-0.12	0.03
Total Effect	0.03	-0.04	0.09	0.00	-0.08	0.07

Note. Estimates under the sequential ignorability and homogeneous interaction assumptions.

Table C7: Estimates for multiple mediation on student outcomes –socioemotional development.

	Follow-up I (June 2016)			Follow-up II (June 2017)		
	Estimate	95% CI Lower	95% CI Upper	Estimate	95% CI Lower	95% CI Upper
ACME (treated)	0.00	-0.04	0.04	0.00	-0.03	0.04
ACME (control)	0.02	-0.07	0.10	0.02	-0.03	0.06
ACME (average)	0.00	-0.04	0.04	0.01	-0.02	0.04
ADE (treated)	0.21	0.09	0.32	-0.04	-0.14	0.06
ADE (control)	0.23	0.14	0.32	-0.03	-0.12	0.07
ADE (average)	0.21	0.11	0.31	-0.03	-0.13	0.06
Total Effect	0.22	0.14	0.31	-0.02	-0.12	0.07

Note. Estimates under the sequential ignorability and homogeneous interaction assumptions.

Table C8: Estimates for multiple mediation on student outcomes –executive functioning.

	Follow-up I (June 2016)			Follow-up II (June 2017)		
	Estimate	95% CI Lower	95% CI Upper	Estimate	95% CI Lower	95% CI Upper
ACME (treated)	-0.07	-0.11	-0.03	0.01	-0.02	0.04
ACME (control)	-0.06	-0.16	0.03	0.02	-0.02	0.07
ACME (average)	**-0.07**	**-0.11**	**-0.03**	0.02	-0.01	0.04
ADE (treated)	0.17	0.05	0.30	-0.10	-0.21	0.00
ADE (control)	0.18	0.09	0.27	-0.09	-0.19	0.00
ADE (average)	0.18	0.07	0.28	-0.10	-0.20	0.00
Total Effect	0.11	0.03	0.19	-0.08	-0.17	0.01

Note. Estimates under the sequential ignorability and homogeneous interaction assumptions.

REFERENCES

Abadzi, H. (2012). Effective teacher training in low-income countries: the power of observational learning research.

Alterman, A. I., McDermott, P. A., Cacciola, J. S., Rutherford, M. J., Boardman, C. R., McKay, J. R., & Cook, T. G. (1998). A typology of antisociality in methadone patients. *Journal of Abnormal Psychology*, *107*(3), 412.

Ball, D. L., Thames, M. H., & Phelps, G. (2005). Articulating domains of mathematical knowledge for teaching. Retrieved May 13, 2005, from http://wwwpersonal.umich.edu/~dball/Presentations/RecentPresentations/041405_MKT_AERA.pdf.

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukherji, S., ... & Walton, M. (2016). *Mainstreaming an effective intervention: Evidence from randomized evaluations of "Teaching at the Right Level" in India* (No. w22746). National Bureau of Economic Research.

Baron R.M. & Kenny, D.A. (1986). The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology, 51*(6), 1173–1182.

Bau, N. and Jishnu, D. (2017). The Misallocation of Pay and Productivity in the Public Sector: Evidence from the Labor Market for Teachers. Policy Research Working Paper, 8050. World Bank, Washington, D.C.

Bennell, P., and K. Akyeampong. (2007). *Teacher Motivation in Sub-Saharan Africa and South Asia*. Educational Paper No. 71, DFID: London, UK. Retrieved on January 14, 2019 from https://assets.publishing.service.gov.uk/media/57a08be640f0b652dd000f9a/ResearchingtheIssuesNo71.pdf.

Berkel, A. Mauricio, E. Schoenfelder, I. Sandler. Putting the pieces together: an integrated model of program implementation**.** *Prevention Science, 12* (2011), pp. 22-33.

Blömeke, S. & Delaney, S. (2012). Assessment of teacher knowledge across countries: A review of the state of research. *ZDM Mathematics Education, 44*, 223-247.

Bold, T., Filmer, D., Martin, G., Molina, E., Rockmore, C., Stacy, B., Svensson, J., & Wane, W. (2017). What do teachers know and do? does it matter? Evidence from primary schools in Africa. Washington, DC: The World Bank.

Brophy, J. E., & Evertson, C. M. (1976). *Learning from teaching: A developmental perspective*. Allyn and Bacon.

Bradley, R., Danielson, L., & Doolittle, J. (2005). Response to intervention. *Journal of Learning Disabilities*, *38*(6), 485-486

Buhl-Wiggers, J., Kerwin, J. T., Smith, J. A., and Thorton, R. (2017). The impact of teacher effectiveness on student learning in Africa. Research on Improving Systems of Education (RISE) Working Paper.

Burns, M. K., & Gibbons, K. (2013). *Implementing response-to-intervention in elementary and secondary schools: Procedures to assure scientific-based practices*. Routledge.

Burgess, T. (2009). Teacher knowledge and statistics: What types of knowledge are used in the primary classroom?. *The Mathematics Enthusiast*, *6*(1), 3-24.

Cambridge Education. (n.d.). *Teaching in Ghana: Turning the barrier into the solution*.

Carlisle, J. F., Cortina, K. S., & Katz, L. A. (2011). First-grade teachers' response to three models of professional development in reading. *Reading & Writing Quarterly*, *27*(3), 212-238.

Caspary, K. (2002). California's pioneering training and retention initiatives for early childhood educators. *Berkeley, CA: PACE*.

Cattell, R. B. (1949). r_p and other coefficients of pattern similarity. Psychometrika, 14, 279–298.

Chambers, J. G., Lam, I., & Mahitivanichcha, K. (2008). *Examining context and challenges in measuring investment in professional development: A case study of six school districts in the Southwest Region* (Issues & Answers Report, REL 2008–No. 037). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved on October 20, 2019, from http://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/REL_2008037.pdf

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review, 104*(9): 2633-2679.

Cilliers, J., Fleisch, B., Prinsloo, C., & Taylor, S. (2019). How to improve teaching practice? An experimental comparison of centralized training and in-classroom coaching. *Journal of Human Resources*, 0618-9538R1.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). Teacher credentials and student achievement in high school a cross-subject analysis with student fixed effects. *Journal of Human Resources*, *45*(3), 655-681.

Cochran-Smith, M., & Lytle, S. L. (1990). Research on teaching and teacher research: The issues that divide. *Educational researcher*, *19*(2), 2-11.

Coker, H., Medley, D. M., & Soar, R. S. (1980). How valid are expert opinions about effective teaching?. *The Phi Delta Kappan*, *62*(2), 131-149.

Collie, R. J., & Martin, A. J. (2017). Adaptive and maladaptive work-related motivation among teachers: A person-centered examination and links with well-being. *Teaching and Teacher Education*, *64*, 199-210.

Conn, K. M. (2017). Identifying effective education interventions in sub-Saharan Africa: A meta-analysis of impact evaluations. *Review of Educational Research, 87*(5), 863–898.

Connor, C. M., Son, S., Hindman, A. H., & Morrison, F. J. (2005). Teacher qualifications, classroom practices, family characteristics, and preschool experience: Complex effects on first graders' vocabulary and early reading outcomes. *Journal of School Psychology, 43*(4), 343–375.

Crisci, A., Sepe, E., & Malafronte, P. (2019). What influences teachers' job satisfaction and how to improve, develop and reorganize the school activities associated with them. *Quality & Quantity*, *53*(5), 2403-2419.

Danielson, C. (2013). The framework for teaching evaluation instrument, 2013 instructionally focused edition Princeton, NJ: The Danielson Group.

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). Effective teacher professional development.

Das, J., Dercon, S., Habyarimana, J., & Krishnan, P. (2007). Teacher shocks and student learning evidence from Zambia. *Journal of Human Resources*, *42*(4), 820-862.

Desimone L. M., Garet M. S. (2015). Best practices in teachers' professional development in the United States. *Psychology, Society and Education, 7*(3), 252–263.

Dobbie, W., & Fryer Jr, R. G. (2013). Getting beneath the veil of effective schools: Evidence from New York City. *American Economic Journal: Applied Economics*, *5*(4), 28-60.

Domitrovich, C. E., & Greenberg, M. T. (2004). Introduction to the special issue: Preventive interventions with young children: Building on the foundation of early intervention programs. *Early Education & Development*, *15*(4), 365-370.

Dubeck, M. M., Jukes, M. C. H., Brooker, S. J., Drake, T. L., & Inyega, H. N. (2015). Designing a program of teacher professional development to support beginning reading acquisition in coastal Kenya. *International Journal of Educational Development, 41*, 88–96. https//doi.org/10.1016/j.ijedudev.2014.11.022.

Duda, R. O., & Hart, P. E. (1973). *Pattern classification and scene analysis* (Vol. 3). New York: Wiley.

Durlack, J. A. & DuPre, E. P. (2008). Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation**.** *American Journal of Community Psychology, 41* (2008), pp. 327-350.

Edinger, S. K., & Edinger, M. J. (2018). Improving teacher job satisfaction: The roles of social capital, teacher efficacy, and support. *The Journal of psychology*, *152*(8), 573-593.

Egert, F., Fukkink, R. G., & Eckhardt, A. G. (2018). Impact of in-service professional development programs for early childhood teachers on quality ratings and child outcomes: A meta-analysis. *Review of Educational Research, 88*(3), 401–433.

Fatima, S.F. & Wolf, S. (accepted with minor revisions). Cumulative Risk and Newly Qualified Teachers' Professional Well-being: Evidence from Rural Ghana. *Research in Comparative and International Education.*

Ferlazzo, L. (2018, June 8). *Response: Do Professional Development 'With' Teachers, Not 'to' Them.* Retrieved March 1, 2020, from http://blogs.edweek.org/teachers/classroom_qa_with_larry_ferlazzo/2018/06/response_do_professional_development_with_teachers_not_to_them.html

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it?. *Reading research quarterly*, *41*(1), 93-99.

Fisher, C. W., Berliner, D. C., Filby, N. N., Marliave, R., Cahen, L. S., & Dishaw, M. M. (1981). Teaching behaviors, academic learning time, and student achievement: An overview. *The Journal of classroom interaction*, *17*(1), 2-15.

Gage, N. L., & Needels, M. C. (1989). Process-product research on teaching: A review of criticisms. *The elementary school journal*, *89*(3), 253-300.

Garet, M. S., Heppen, J., Walters, K., Smith, T., & Yang, R. (2016). Does content-focused teacher professional development work? Findings from three Institute of Education Sciences studies. Retrieved from the Institute of Education Sciences.

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

Greenberg, M. T., Brown, J. L., & Abenavoli, R. M. (2016). *Teacher stress and health effects on teachers, students, and schools.* Edna Bennett Pierce Prevention Research Center, Pennsylvania State University. Retrieved on January 25, 2019 from http://prevention.psu.edu/uploads/files/rwjf430428.pdf

Gregson, J. A., & Sturko, P. A. (2007). Teachers as adult learners: Re-conceptualizing professional development. *Journal of adult education*, *36*(1), 1-18.

Grossman, P., Loeb, S., Cohen, J., & Wyckoff, J. (2013). Measure for measure: The relationship between measures of instructional practice in middle school English language arts and teachers' value-added scores. *American Journal of Education. 119*, 445-470.

Guerriero, S. (2014). Teachers' pedagogical knowledge and the teaching profession. *Teaching and Teacher Education*, *2*(1), 7.

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and teaching*, *8*(3), 381-391.

Guskey, T. R., & Yoon, K. S. (2009). What works in professional development?. *Phi delta kappan*, *90*(7), 495-500.

Hallman-Thrasher, A., Connor, J., & Sturgill, D. (2019). Strong discipline knowledge cuts both ways for novice mathematics and science teachers. *International Journal of Science and Mathematics Education*, *17*(2), 253-272.

Halpin, P. F. & Kieffer, M. J. (2015). Describing profiles of instructional practice: A new approach to analyzing classroom observation data. *Educational Researcher, 44*, 263-277.

Hamre, B. K. (2014). Teachers' daily interactions with children: An essential ingredient in effective early childhood programs. *Child Development Perspectives*, *8*(4), 223-230.

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

Hamre, B. K., Partee, A., & Mulcahy, C. (2017). Enhancing the impact of professional development in the context of preschool expansion.

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. Routledge.

Ho, A. D., & Kane, T. J. (2012). The reliability of classroom observations by school personnel. Bill and Melinda Gates Foundation.

Hulleman, C. S. & Cordray, D. (2009). Moving from the lab to the field: the role of fidelity and achieved relative intervention strength**.** *Journal of Research on Educational Effectiveness, 2*, pp. 88-110.

Imai, K., Keele, L., & Tingley, D. (2010). "A General Approach to Causal Mediation Analysis." *Psychological Methods, 15*(4), 309–334.

Imai, K., Keele, L., Tingley, D., & Yamamoto, T. (2011). "Unpacking the Black Box: Learning about Causal Mechanisms from Experimental and Observational Studies." *American Political Science Review, 105*(4), 765–789.

Imai, K., & Yamamoto, T. (2013). Identification and sensitivity analysis for multiple causal mechanisms: Revisiting evidence from framing experiments. *Political Analysis*, *21*(2), 141-171.

Jimerson, S. R., Burns, M. K., & VanDerHeyden, A. M. (2007). Response to intervention at school: The science and practice of assessment and intervention. In *Handbook of response to intervention* (pp. 3-9). Springer, Boston, MA.

Jukes, M. C., Turner, E. L., Dubeck, M. M., Halliday, K. E., Inyega, H. N., Wolf, S., Zuilkowski, S.S., & Brooker, S. J. (2017). Improving literacy instruction in Kenya through teacher professional development and text messages support: A cluster randomized trial. *Journal of Research on Educational Effectiveness*, *10*(3), 449-481.

Kabay, S., Wolf, S., & Yoshikawa, H. (2017). “So that his mind will open”: Parental perceptions of preschool in urbanizing Ghana. *International Journal of Educational Development, 57*, 44–53.

Kaffenberger, M. (2019, May 17). What have we learned about the learning crisis? Retrieved October 24, 2019, from https://www.brookings.edu/blog/education-plusdevelopment/2019/05/17/what-have-we-learned-about-the-learning-crisis/.

Kale-Dery, S. (2018, August 2). *Shift education system towards life-long learning*. Retrieved October 24, 2019, from http://www.t-tel.org/news-view/shift-education-system-towards-life-long-learning

Kane, T. J. and D. O. Staiger (2008). Estimating teacher impacts on student achievement: An experimental evaluation. NBER Working Paper #14607. Cambridge, MA.

Kelcey, B., Shen, Z., & Spybrook, J. (2016). Intraclass correlation coefficients for designing cluster-randomized trials in sub-Saharan Africa education. *Evaluation review*, *40*(6), 500-525.

Klassen, R. M., Perry, N. E., & Frenzel, A. C. (2012). Teachers' relatedness with students: An underemphasized component of teachers' basic psychological needs. *Journal of Educational Psychology*, *104*(1), 150.

Klieme, E., Lipowsky, F., Rakoczy, K., & Ratzka, N. (2006). Quality dimensions and effectiveness of mathematics education. *Studies on the educational quality of school*, 127-146.

Klusmann, U., Kunter, M., Trautwein, U., Lüdtke, O., & Baumert, J. (2008). Teachers' occupational well-being and quality of instruction: The important role of self-regulatory patterns. *Journal of Educational Psychology*, *100*(3), 702.

Konold, T. R., Glutting, J. J., McDermott, P. A., Kush, J. C., & Watkins, M. M. (1999). Structure and diagnostic benefits of a normative subtest taxonomy developed from the WISC-III standardization sample. *Journal of School Psychology*, *37*(1), 29-48.

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

Lazarev, V. & Newman, D. (2014). Can Multifactor Models of Teaching Improve Teacher Effectiveness Measures? In. S. Corcoran (Chair), Policy Considerations in the Implementation of MMTES. Paper session conducted at The Association

for Education Finance and Policy, San Antonio, TX. Retrieved on October 20, 2019, from http://ssrn.com/abstract=2493544

Lee, J. E. (2017). Preschool teachers' pedagogical content knowledge in mathematics. *International Journal of Early Childhood*, *49*(2), 229-243.

Leyva, D., Weiland, C., Barata, M., Yoshikawa, H., Snow, C., Treviño, E., & Rolla, A. (2015). Teacher–child interactions in Chile and their associations with prekindergarten outcomes. *Child development*, *86*(3), 781-799.

LoCasale-Crouch, J., DeCoster, J., Cabell, S. Q., Pianta, R. C., Hamre, B. K., Downer, J. T., & LaParo, K. (2016). Unpacking intervention effects: Teacher responsiveness as a mediator of perceived intervention quality and change in teaching practice. *Early childhood research quarterly*, *36*, 201-209.

Martin, A. J. (2009). Motivation and engagement in the workplace: Examining a multidimensional framework and instrument from a measurement and evaluation perspective. *Measurement and Evaluation in Counseling and Development*, *41*(4), 223-243.

Maslach, C., Jackson, S. E., & Leiter, M. P. (1996). *MBI: Maslach burnout inventory*. Sunnyvale (CA): CPP, Incorporated.

Maslach, C., Schaufeli, W. B., & Leiter, M. P. (2001). Job burnout. *Annual Review of Psychology*, *52*(1), 397-422.

Matsumura, L. C., Garnier, H. E., Correnti, R., Junker, B., & Bickel, D. D. (2010). Investigating the effectiveness of a comprehensive literacy coaching program in schools with high teacher mobility. *The Elementary School Journal, 111*(1), 35–62.

McDermott, P. A., Glutting, J. J., Jones, J. N., Watkins, M. W., & Kush, J. (1989). Identification and membership of core profile types in the WISC-R national standardization sample. *Psychological Assessment, 1*, 292–299.

Molina, E., Fatima, S. F., Ho, A., Hurtado, C. M., Wilichowksi, T., & Pushparatnam, A. (2018). Measuring Teaching Practices at Scale: Results from the Development and Validation of the Teach Classroom Observation Tool.

McLean, L., & Connor, C. M. (2015). Depressive symptoms in third□grade teachers: Relations to classroom quality and student achievement. *Child Development*, *86*(3), 945-954.

Ministry of Education, Ghana Education Service. (2012). *Pre-Tertiary Teacher Professional Development and Management in Ghana*.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers college record*, *108*(6), 1017-1054.

Mikesell, D. N. (2018). *Concept Project Information Document (PID) - Ghana Accountability for Learning Outcomes Project - P165557 (English)*. Washington, D.C.: World Bank Group. Retrieved on June 20, 2019 from http://documents.worldbank.org/curated/en/612671545310563888/Concept-Project-Information-Document-PID-Ghana-Accountability-for-Learning-Outcomes-Project-P165557

Moir, E. (2015, September 15). Evolving from professional development to professional learning. Retrieved March 1, 2020, from https://edsource.org/2013/evolving-from-professional-development-to-professional-learning/32586

Muijs, D., Kyriakides, L., Van der Werf, G., Creemers, B., Timperley, H., & Earl, L. (2014). State of the art–teacher effectiveness and professional learning. *School effectiveness and school improvement*, *25*(2), 231-256.

Mulkeen, A. (2010). *Teachers in Anglophone Africa: Issues in teacher supply, training, and management*. Washington: World Bank Publications. Retrieved on January 14, 2019 from http://siteresources.worldbank.org/EDUCATION/Resources/278200-1099079877269/Teachers_Anglophone_Africa.pdf

Myers, D., Scott, T., & Sugai, G. (2014). Multi-tiered support framework for teachers' classroom management practices: Overview and case study of building the triangle for teachers. *Journal of Positive Behavior Interventions, 16*, 179–190. doi:10.1177/1098300713484062

NASUWT [National Association of Schoolmasters Union of Women Teachers] (2017). *The big question 2017: An opinion survey of teachers and school leaders*. Retrieved on January 14, 2019, from www.nasuwt.org.uk/uploads/assets/uploaded/7649b810-30c7-4e93-986b363487926b1d.pdf.

OECD [Organization for Economic Cooperation and Development] (2018). Teachers' well-being, confidence and efficacy in *Valuing our Teachers and Raising their Status: How Communities Can Help*. OECD Publishing, Paris. Retrieved on January 14, 2019 from https://doi.org/10.1787/9789264292697-6-en.

Overall, J. E., & Magee, K. N. (1992). Replication as a rule for determining the number of clusters in hierarchical cluster analysis. *Applied Psychological Measurement*, *16*(2), 119-128.

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

Pianta R.C., Hamre B.K., Hayes N., Mintz S., LaParo K.M. (2008). Classroom Assessment Scoring System–Secondary (CLASS-S) Charlottesville, VA: University of Virginia.

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38*(2), 109–119.

Piper, B., Zuilkowski, S. S., Dubeck, M., Jepkemei, E., & King, S. J. (2018). Identifying the essential ingredients to literacy and numeracy improvement: Teacher professional development and coaching, student textbooks, and structured teachers' guides. *World Development*, *106*, 324-336.

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

Portin, B. S., Alejano, C. R., Knapp, M. S., & Marzolf, E. (2006). Redefining roles, responsibilities, and authority of school leaders. Seattle, WA: Center for the Study of Teaching and Policy.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*(3), 879-891.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, *94*(2), 247-252.

RTI International. (2019). Training teachers or robots: Unexpected findings of a 7 country teacher professional development study. Conference proceedings at Comparative International Education Society. San Francisco, CA.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68.

Salmela-Aro, K., Hietajärvi, L., & Lonka, K. (2019). Work Burnout and Engagement Profiles Among Teachers. *Frontiers in Psychology*, *10*, 2254.

Seidman, E., Raza, M., Kim, S., & McCoy, J. M. (2014). Teacher Instructional Practices and Processes System (V.5) – TIPPS: Manual and scoring system. New York: NYU.

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

Sharma, R. D., & Jyoti, J. (2006). Job satisfaction among school teachers. *IIMB Management Review*, *18*(4), 349-363.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4-14.

Simonsen, B., Fairbanks, S., Briesch, A., Myers, D., & Sugai, G. (2008). A review of evidence based practices in class-room management: Considerations for research to practice. *Education & Treatment of Children, 31*, 351–380.

Simonsen, B., Myers, D., & DeLuca, C. (2010). Providing teachers with training and performance feedback to increase use of three classroom management skills: Prompts, opportunities to respond, and reinforcement. *Teacher Education in Special Education, 33*, 300–318.

Simonsen, B., MacSuga-Gage, A. S., Briere, D. E., Freeman, J., Myers, D., Scott, T., & Sugai, G. (2014). Multi-tiered support framework for educators' classroom management practices: Overview and case study of building the triangle for educators. Journal of Positive Behavior Interventions, 16, 179–190.

Smith, P. L., Waller, M. I., & Waller, S. P. (1982). Generalizable observation of the teaching process. *Educational and Psychological Measurement*, *42*(2), 467-478.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equations models. In S. Leinhart (Ed.), Sociological methodology 1982 (pp. 290-312). San Francisco: Jossey-Bass.

Sobel, M. E. (1986). Some new results on indirect effects and their standard errors in covariance structure models. In N. Tuma (Ed.), Sociological Methodology 1986 (pp. 159-186). Washington, DC: American Sociological Association.

Stallings, J., Almy, M., Resnick, L. B., & Leinhardt, G. (1975). Implementation and child effects of teaching practices in Follow Through classrooms. *Monographs of the society for research in child development*, 1-133.

Swaner, L. (2016). Teachers as adult learners: Implications of theory and research for designing effective professional development. CSE. Volume 20. Number 3.43. Retrieved March 1, 2020 from www.acsi.org/documents/resources/CSE/teachers-as-adult-learners

Teemant, A. (2014). A mixed-methods investigation of instructional coaching for teachers of diverse learners. *Urban Education*, *49*(5), 574-604.

Thames, M. H., & Ball, D. L. (2010). What Math Knowledge Does Teaching Require?. *Teaching Children Mathematics*, *17*(4), 220-229.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package for causal mediation analysis.

Tryon, R. C. & Bailey, D. E. (1970). Cluster analysis. New York: McGraw-Hill.

UNDP [United Nations Development Program]. (2018). Human development indices and indicators: 2018 Statistical update.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *American Statistical Association Journal, 58*, 236–244.

WGU [Western Governors University]. (2019, August 2). Professional Development vs Professional Learning. Retrieved March 1, 2020, from https://www.wgu.edu/blog/professional-development-vs-professional-learning-teachers1712.html

Wolf, S. (2018). Impacts of pre-service training and coaching on kindergarten quality and student learning outcomes in Ghana. *Studies in Educational Evaluation*, *59*, 112-123.

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

Wolf, S., Aber, J.L., Behrman, J.R. & Tsinigo, E. (2018a). Experimental Impacts of the Quality Preschool for Ghana Interventions on Teacher Professional Well-being, Classroom Quality, and Children's School Readiness, *Journal of Research on Educational Effectiveness, 12*(1), 10-37.

Wolf, S., Raza, M., Kim, S., Aber, J. L., Behrman, J., & Seidman, E. (2018b). Measuring and predicting process quality in Ghanaian pre-primary classrooms using the Teacher Instructional Practices and Processes System (TIPPS). *Early Childhood Research Quarterly*, *45*, 18-30.

World Bank. (2018). World Development Report 2018: Learning to Realize Education's Promise. Washington, DC: World Bank.

World Bank. (2019, January 22). The Education Crisis: Being in School Is Not the Same as Learning. Retrieved October 24, 2019, from https://www.worldbank.org/en/news/immersive-story/2019/01/22/pass-or-fail-how-can-the-world-do-its-homework.