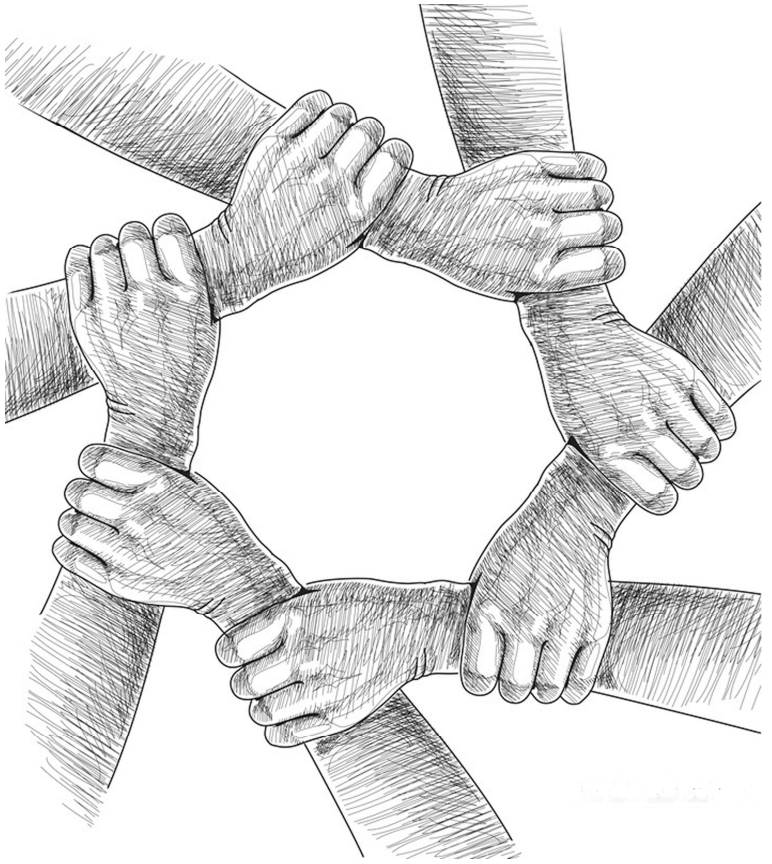
The Influence of Within School and Across Schools' Collaborative Practices on Student Learning and Teaching Outcomes in West Africa



Fernanda Soares

The influence of within school and across schools' collaborative practices on student learning and teaching outcomes in West Africa.

ISBN 978 90 8666 498 6
Publisher: Boekenplan, Maastricht
www.boekenplan.nl

The influence of within school and across schools' collaborative practices on student learning and teaching outcomes in West Africa.

DISSERTATION

to obtain the degree of Doctor at
Maastricht University, on the authority of the Rector Magnificus Prof. Dr.
Rianne M. Letschert.
in accordance with the decision of the Board of Deans,
to be defended in public on
Monday November 18, 2019 at 16.00 hrs.

By

Fernanda Soares

Supervisors

Prof. Dr. Kristof De Witte

Prof. Dr. J. Dronkers †

Co-Supervisor

Dr. Mindel van de Laar

Assessment Committee

Prof. Dr. Wim Groot (Chairman)

Prof. Dr. Jo Ritzen

Prof. Dr. Peter Geschiere, Universiteit van Amsterdam

Prof. Dr. Wouter Schelfhout, Universiteit van Antwerpen

Acknowledgments

I would like to thank some of the many people who provided support and guidance throughout my PhD journey.

First, I would like to express my greatest gratitude to Prof. Dronkers, Prof. De Witte and Dr. van de Laar. My sincere gratitude to Prof. Dronkers for his mentoring, wisdom and interest on this research. I thank Prof. de Witte for providing great tutelage and for always encouraging me to think outside the box and to apply rigorous methods. I thank Dr. van de Laar for her steady and continuous support from day one as GPAC coordinator and subsequently for her guidance as my supervisor. I was extremely lucky to have this outstanding team committed to this research and continuously supporting me throughout the way.

I would also like to express my gratitude to the assessment committee members Prof. Dr. Wim Groot, Prof. Dr. Jo Ritzen, Prof. Dr. Peter Geschiere and Prof. Dr. Wouter Schelfhout for generously allocating time to carefully read this dissertation and provide valuable recommendations.

I sincerely maintain a high level of appreciation for many people from UNU-MERIT who enabled me to complete this research. I would like to thank the GPAC^2 coordination team (Charlotte, Shivani, Saba, Carlos, Cristina) for their support during GPAC^2 seminars. I would also like to express my gratitude to many professors and students from UNU-MERIT who provided valuable comments to my research during GPAC^2 seminars. I also would like to thank wonderful colleagues I have met and that were part this journey: Michal, Vincenzo, Shea, Karen, Hana, Felipe and Abel.

I am very grateful for the support that I have received from my current employer, Family Health International 360 (FHI360). I would also like to thank my previous academic mentors, Prof. Matilde de Souza, Prof. Paul Winters, and Dr. Suzanne Duryea, who have contributed to my development as a researcher throughout my academic and professional trajectory.

Finally, I would like to dedicate this thesis to my family. To my mother, without her I literally would not be here. Beyond that, she taught me, from a very young age, that the only asset you acquire in life that no one can take away from you is education. She supported me in obtaining the education I needed to embark on a PhD and instilled in me that in life a number of non-cognitive skills are as valuable (if not more valuable) than cognitive skills, which were key for me to succeed in this PhD journey. To my husband, for always believing in me (even when I didn't) and encouraging me to pursue my dreams; for his comprehension, support and unending understanding throughout the PhD studies. To my father, who sadly departed in the middle of this journey, but who was my biggest cheerleader and who would be so proud of me. To my brother, who is always by my side and has a way of helping me see the lighter side of things. To my grandparents, for their love, and to my

daughter Julia, who joined our family in the middle of this PhD journey and who inspires me every day to be a better person and a role model for her.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1. General Introduction

Many low- and middle-income countries have made considerable progress in the past decades in expanding enrollment and in increasing years of schooling (World Bank, 2018; UIS, 2016). The strongest schooling expansions have occurred at the primary level. In Sub-Saharan Africa, gross primary enrollment rate[1] went from 47 percent in 1970 to above 100 percent[2] by 2010 (World Bank, 2018). According to World Bank data, net primary enrollment rate[3] went from 44 percent to almost 78 percent in 2017.[4] Because in Sub-Saharan Africa many children enter primary school late, over-age enrollment, which refer those enrolled who are older than the official school-age range for primary education, is the main reason behind the differences between gross primary enrollment rate and net primary enrollment rate (Lewin, 2009).[5]

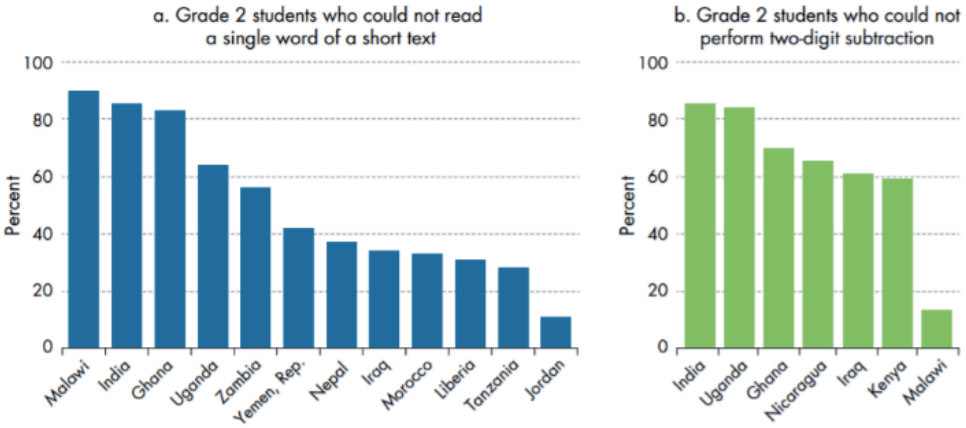
The expansion in schooling has been accompanied by a learning crisis, characterized by a higher proportion of children attending school, but failing to learn the expected competences for their grades (UNESCO, 2013). Although the learning crisis is global, it concentrates in low-and middle-income countries. A large share of children in low-income countries complete their primary education lacking basic reading, writing, and arithmetic skills (Bold et al., 2017). In Sub-Saharan Africa, in particular, regional and international learning assessments show that less than 50 percent of the students tested reached the absolute minimum level of learning, as defined by the various assessments (Bashir et al., 2018). The learning crisis emerges already in the early grades. Early-grade reading and mathematics assessments at the end of grade 2 reveal that over 80 percent of students in Ghana and Malawi were unable to read a single word of a short text while over 60 percent of students in in Uganda and Ghana could not perform two-digit subtraction (Figure 1.1).

Even when second grade children in Sub-Saharan are able to read a short passage, the majority (50 to 80 percent) cannot answer a single question based the passage (Bashir et al., 2018). Figure 1.2 illustrates results from Early Grade Reading Assessments (EGRA) applied across Sub-Saharan countries showing that a high share of second-grade (and third-grade) students receives scores of zero on assessments of reading comprehension. This indicates that children in early grades lack the foundation literacy skills to transition from "learning to read" to "reading to learn". Indeed, the learning deficits from early grades are carried over to subsequent years, culminating in a high

[1] The gross primary enrollment ratio is defined by the UNESCO Institute for Statistics (UIS) (2018) as the number of students enrolled in a primary education, regardless of age, expressed as a percentage of the official school-age population corresponding to primary education level.
[2] Because primary gross enrollment includes students whose age exceeds the official age group for primary education level, the rate may exceed 100 percent.
[3] The net primary enrollment ratio is defined by the UNESCO Institute for Statistics (UIS) (2018) as the number of age appropriate students enrolled in a primary education education, expressed as a percentage of the official school-age population corresponding to primary education level.
[4] Word Bank site: https://data.worldbank.org/indicator/SE.PRM.NENR?locations=ZG?. Accessed on June 16th, 2019.
[5] Although there has been considerable expansion it is important to note that in Sub-Saharan Africa about 21% of children of primary school age out of school (UIS, 2016).

percentage of students who complete primary education without mastering basic competences. According to PASEC 2014, which is a regional learning assessment in West and Central Africa, less than 45 percent of grade six students reached the "sufficient" competency level for continuing studies in reading or mathematics (Malpel, 2016). Similarly, results from the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) 2007, which is applied in southern and East Africa, reveal that 37 percent of grade six students are not competent in reading, and more than 60 percent are not competent in mathematics (World Bank, 2018).

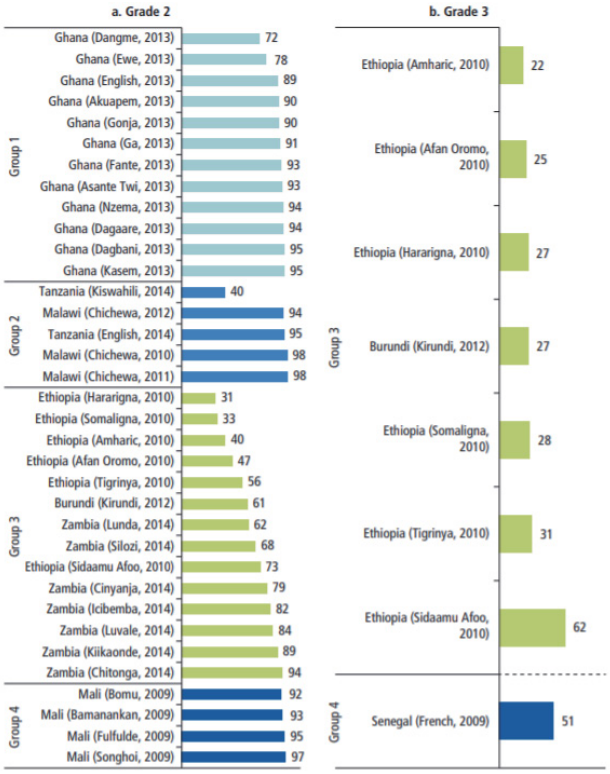
Figure 1.1 Percentage of grade 2 students who could not perform simple reading or math tasks, selected countries



Notes: Figure uses reading and mathematics data for Kenya and Uganda from Uwezo Annual Assessment Reports, 2015; reading and mathematics data for rural India from ASER Centre (2017); reading data for all other countries from U.S. Agency for International Development (USAID), Early Grade Reading Barometer, 2017; and mathematics data for all other countries from USAID/RTI Early Grade Mathematics Assessment intervention reports, 2012–15. These data pertain to selected regions in the countries and are not nationally representative.
Source: World Bank, 2018, p.5.

The World Bank (2018) compiled data from learning assessments in 95 countries and established a globally comparable "minimum proficiency" threshold. Falling below this threshold in math means that students have not mastered basic mathematical skills, such as making simple computations with whole numbers, using fractions or measurements, or interpreting simple bar graphs. For reading, minimum proficiency means that students can locate and retrieve explicitly stated detail when reading literary texts and can locate and reproduce explicitly stated information from the beginning of informational texts. Results illustrated in Figure 1.3 show that more than 60 percent of primary school children in low- and middle- income countries fail to achieve minimum proficiency in math and reading. Results disaggregated by region show that the situation is even more acute in Sub-Saharan African countries, where over 80 percent of students do not achieve minimum proficiency.

Figure 1.2 Percentage of Students in Grade Two and Three Receiving EGRA Reading Comprehension Scores of Zero in Selected Sub-Saharan African Countries, by Testing Language and Country Group

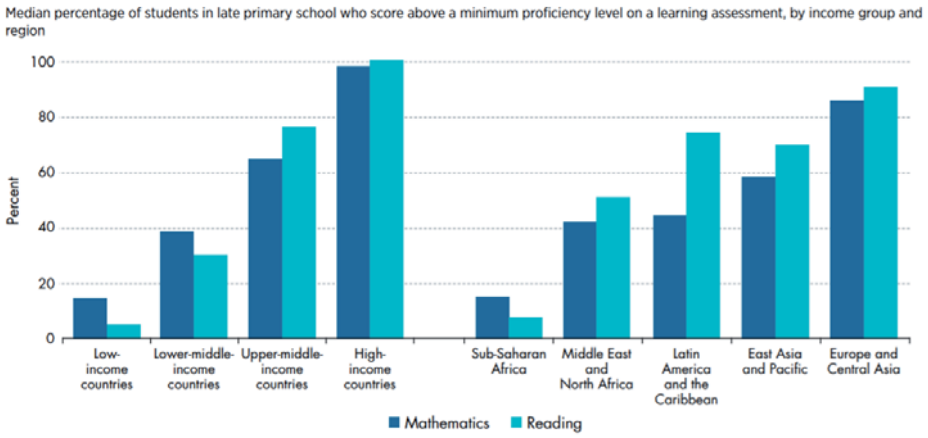


Note: Figure uses data from Early Grade Reading Assessment (EGRA) World Bank EdStats database (accessed January 4, 2017). After each country name, the language and year in which the test was administered appears within parentheses.

Source: Bashir et al., 2018, p. 79

Research in francophone Sub-Saharan Africa has shown that students' family background is one of the largest overall predictor of student learning outcome (Michaelowa, 2001). However, research over the past decade, especially in Western countries, has also built new evidence that once children get to school, no single factor is as critical as the quality of teachers and teaching practices (Bruns & Luque, 2014). In the sections that follow, we provide a brief overview of the literature on teachers' influence on student achievement and on how to improve teaching quality through in-service professional development, with a focus on peer collaboration and in-school pedagogical support.

Figure 1.3. The percentage of primary school students who pass a minimum proficiency threshold is often low



Notes: Bars show the unweighted cross-country median within country grouping. Regional averages exclude high-income countries. India and China are among the countries excluded for lack of data. Minimum proficiency in mathematics is benchmarked to the Trends in International Mathematics and Science Study (TIMSS) assessment and in reading to the Progress in International Reading Literacy Study (PIRLS) assessment. Minimum proficiency in mathematics means that students have some basic mathematical knowledge such as adding or subtracting whole numbers, recognizing familiar geometric shapes, and reading simple graphs and tables (Mullis and others 2016). Minimum proficiency in reading means that students can locate and retrieve explicitly stated detail when reading literary texts and can locate and reproduce explicitly stated information from the beginning of informational texts (Mullis and others 2012).

Source: World Bank, 2018, p. 8.

1.1. The role of teachers

There is a broad literature from developed countries suggesting that teacher quality, broadly defined, matter for student learning (Chetty et al., 2014a; Rivkin, et al., 2005; Rockoff, 2004; Harris & Sass 2006; Kane et al., 2008; Buddin and Zamarro 2009). Past literature that investigates teacher quality has adopted two main lines of research. The first line investigates the extent to which specific teacher characteristics account for differences in student achievement. These studies use the education production function approach to link specific teacher's characteristics (input) to student's outcomes, controlling for student differences (Aslam & Kingdon, 2011). Several of this studies undertake in the United States have found that teachers' expertise (including scores on examinations measuring skills and knowledge), educational level, licensing and experience are associated with student achievement (Fuller, 1999; Ferguson 1991; Strauss & Sawyer, 1986). Also, in Brazil, Harbison and Hanushek (1992) found that teacher education had a significantly positive impact for math on primary school students.

Initial studies employing the education production function relied on cross-section data and lacked controls for student prior achievements. One main criticism to this approach relates to the problem of omitted variable bias, since there are numerous past and current factors that affect achievement (Hanushek & Rivkin, 2006). Recent researches have adopted a value-added modeling approach that

aims to mitigate problems of omitted variable bias. In value added models, achievement scores are a function of current school and family inputs with student fixed effects to control for prior inputs (Buddin & Zamarro, 2009). Though this approach accounts for prior achievements, it may still not capture current factors that affect student's achievement (Hanushek & Rivkin, 2006).

A second line of research drops the parametric, inputs based approach and attempts to identify the total impact of teacher on student learning without relying on specific measurable teachers' characteristics (Hanushek & Rivkin, 2006). The approach calculates teacher quality as a teacher fixed effect in an equation of student achievement gain where different groups of students (in a given year or over time) are taught by the same teacher (Aslam & Kingdon, 2011). Under this value-added gains modeling approach, a teacher who consistently produces high achievement growth is considered a good teacher; alternatively, a teacher who consistently produces low achievement growth is considered a poor teacher (Hanushek & Rivkin, 2006). Studies that adopt the value-added gains approach consistently find that teacher quality has large effects on student achievement (Kane et al., 2008; Aaronson el al. 2007; Rivkin et al., 2005; Rockoff, 2004). Recent value-added gains research in the United States also shown teachers who raise student achievement also improve long-term outcomes, such as likelihood to attend college, earn higher salaries and be less likely to have children as teenagers. However, the results from both lines of research (value added and value added gains) show that observable specific teacher characteristics, such as teacher experience, certification and education, explain little or nothing of the variation in student achievement (Hanushek & Rivkin, 2006).

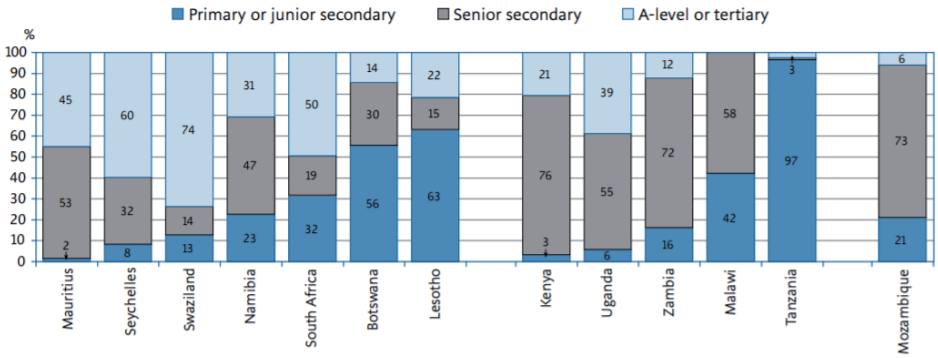
Most of the available research examines evidence from schools in the United States, but related evidence is becoming available for developing countries. Hanushek (1995) reviews the literature available until 1995 and conclude that there is considerable similarity in the research findings across developed and developing countries, except that there is slightly stronger support for a positive impact of teacher education in developing countries. Given the low levels of pre-service and in-service teacher training in developing countries, it is not surprising that in these contexts more education for teachers has a potential to improve students performance. Glewwe & Kremer (2006) reviewed four studies published since the mid-1990s focusing on developing countries[6]. The authors find that teacher variables impact differed widely across the four studies and a systematic relationship with student's achievement could not be observed. Using data from the province of Punjab, Pakistan, Bau and Dass (2017) find a large effect of good teachers on student test scores: moving a student from the 5th to the 95th percentile in the public-school teacher value added distribution would increase mean student test scores by 0.54 standard deviations. Analyzing data

[6] Studies reviewd by Glewwe & Kremer (2006) include: the research of Ghanaian middle schools by Glewwe and Jacoby (1994); the study of Jamaican primary schools by Glewwe et al. (1995); the investigation of grade 8 students in India by Kingdon (1996); and the paper on Philippines primary schools by Tan, Lane and Coustere (1997).

from seven Sub-Sahara African countries, Bold and colleagues (2017) also find significant and large positive effects of teacher content and pedagogical knowledge on student achievement. Finally, Araujo et al. (2016) find that an increase in classroom quality in Ecuador results in higher test scores in language, math, and executive function for kindergarten students.

Despite the recognized importance of teachers to student learning, many teachers in low- and middle-income countries often have low educational qualifications. Data gathered by SACMEQ on the education and training levels of 6th grade teachers show great variation across the 13 countries that participated in the study between 2000 and 2002 (see Figure 1.4). Nonetheless, with a few exceptions (such as Seychelles, Swaziland and South Africa), in most countries the majority of the students are taught reading by teachers holding senior secondary qualification or less (UIS, 2006). In some countries like Tanzania, Lesotho and Botswana the situation is even more worrisome, as 97, 63, and 56 percent of students are taught reading by teachers holding a lower secondary qualification or less (UIS, 2006).

Figure 1.4. Percentage of 6th grade students with reading teachers having completed various education levels, 2000-2002



Notes: Figure uses data from the Southern and Eastern Africa Consortium for Monitoring Education Quality (SACMEQ), 2000-2002. In this figure, as in others based on SACMEQ data, countries are grouped according to need for expansion of primary teacher stock by 2015. The highest expected need for teachers is found in Mozambique.

Source: UIS, 2006, p. 58.

In addition to low-levels of education, teachers in low- and middle-income often lack the subject knowledge, pedagogical skills or motivation to be effective teachers. SACMEQ tested 6th grade reading and mathematics teachers in their respective fields of instruction. Their average scores were then compared to those of their students. Results indicated that in many countries, the highest performing students scored similar levels to that of the average teacher in their country (UIS, 2006). Results also showed a positive relationship between 6th grade teachers' educational levels and their academic skills in reading and math (UIS, 2006). Similarly, Bold and colleagues (2017) analyze data from a nationally representative survey of teachers applied in seven Sub-Saharan African

countries. The authors find that in all countries in the study a large share of teachers do not master the curricula; basic pedagogical knowledge is low; and the use of good teaching practices is rare. Overall, the lack of teaching quality in low-and middle-income countries generally, and in Sub-Saharan African in particularly, may help explain the low levels of student learning in early grade and why so many students complete their primary education lacking even basic reading, writing, and arithmetic skills (Bold et al., 2017).

Pre-service teacher education - *i.e.,* the education teachers receive before they enter into service as a teacher - plays an important role in addressing the problems of teachers' low-levels of education and lack of appropriate knowledge and skills. However, the supply of graduates from pre-service teacher education programs are often not sufficient to attend the demand for qualified teachers in any one country, given the growth in school enrollment and number of schools being established (Junaid & Maka, 2015). As results, uncertified and unprepared teachers are often recruited to fill teaching vacancies in Sub-Saharan African countries that would otherwise remain unfilled (Junaid & Maka, 2015). With the objective of raising teaching quality in Sub-Saharan African countries, there has been a significant increase in opportunities for in-service teacher professional development, broadly understood as any learning opportunity for practicing teachers. In-service training of teachers in Sub-Saharan Africa take many different forms, ranging from one-off trainings of a short duration at designated centers using the cascade model (training of trainers at the central level that in turn deliver trainings at the local level), to more extensive school-based professional development involving peer learning and educator mentors and inspectors visiting classrooms to observe and mentor teachers over a long period of time (Junaid & Maka, 2015). Given the growth in in-service teacher education in the region and the range of mechanisms through which they are delivered, a much greater effort is required to understand its effectiveness and improve its delivery. The following section provides an overview of the role played by in-service professional development initiatives in raising teaching quality and promoting student achievement.

1.2. The Role of in-service professional development in raising teaching quality

Improving teaching through in-service teacher professional development has been identified by systematic reviews of impact evaluation studies as a key for improving student learning in primary schools in low-and middle-income countries (Evans & Popova, 2015; Snilstveit, 2015; Conn, 2014). In a meta-analysis of learning interventions in Sub-Saharan Africa, Conn (2014) finds that pedagogical interventions are more effective at improving student learning than all the other eleven intervention types in her sample (e.g., school management programs, school supplies interventions, or interventions that change the class size or composition). Conn (2014) defines pedagogical interventions as any intervention that aims to change methods of instruction and learning, such as one-time in-service trainings at a central location, long-term teacher mentoring and in-school teacher

coaching. A review of interventions for improving learning outcomes in low-and-middle income countries by Snilstveit (2015) has found that structured pedagogy programs (which are defined by the authors as interventions including development of new content focused on a particular topic, materials for students and teachers, and short term training courses for teachers in delivering the new content) have the largest and most consistent positive average effects on learning outcomes.

The available evidence suggests that in-service professional development can indeed improve student learning. However, many evaluation studies fail to provide details on the content or delivery mechanisms of the interventions that can inform the design of successful programs (Evans & Popova, 2015). Bruns and Lucque (2014) propose a categorization of teacher in-service professional development that sheds some light on the discussion around delivery mechanisms. The authors propose the following four categories: (i) *scripted approaches*, which refer to trainings that prepare teachers to use specific teaching strategies and accompanying materials in the delivery of a well-defined daily curriculum; (ii) *content mastery*, which refer to trainings focused on deepening teachers' expertise in their subject area; (iii) *classroom management,* which are trainings focused on improving teachers' classroom effectiveness; and (iv) *peer collaboration,* which are school-based or cross-school structured opportunities for groups of teachers learn from each other's practice and collaborate on a number of areas, such as curriculum development, student assessment, research and other activities. The category of peer collaboration proposed also includes school-based instructional coaching, provided by principals, school inspectors or other educational actors. In the next section, we focus on *peer collaboration* as a category of in-service professional development proposed by Bruns and Lucque (2014), which is often under-studied in developing countries and is the focus of this dissertation.

1.3. Peer collaboration and school-based pedagogical support

While there is some research from developing countries on the impact of teacher training programs on student learning (examples include Lucas et al., 2014; Chesterfield & Abreu-Combs, 2011; Sailors et al., 2010; Piper & Korda, 2010; Piper, 2009), the majority of quantitative research on peer collaboration and in-school teacher pedagogical support has been conducted in Western countries.[7]

[7] Exceptions include a few studies on coaching. Research conducted in Kenya and Thailand suggested that school-based instructional coaching can lead to teacher behavior change (Piper et al., 2014) and to greater use of student-centered pedagogies (Tolleyet al., 2012). In randomized evaluation of a coaching program in the Brazilian state of Ceará, Bruns et al. (2018) found that the coaching increased teachers' time on instruction and student engagement and produced statistically significant gains in student learning. Another exception is the study by Hardman et al. (2009), which investigates the impact of a school-based teacher development program based on coaching and peer collaboration on learning and teaching in Kenyan primary schools. The authors find that compared to the earlier baseline, teachers in the program were more interactive with the pupils in their whole class teaching and used more group work. The greatest impact on classroom practice was seen in the classrooms of those teachers who had received direct in-service training and were responsible for training their colleagues in the school to pass on their training. However, the cascade training model was relatively ineffective; non-directly trained teachers did not change their teaching as much as the directly

In this section we provide a brief overview of the literature on Professional Learning Communities (PLCs) as form of peer collaboration as well as on the role of principals in providing pedagogical support to teachers and more broadly in raising student achievement.

Professional Learning Communities

There is no broad consensus in the literature on a definition of PLC (Lomos, 2011; Vescio, Ross, & Adams, 2007; Stoll et al., 2006). However, most definitions seem to agree that PLCs involve a group of teachers sharing and critically interrogating their practice in an ongoing, reflective, collaborative, inclusive, learning-oriented, growth-promoting way (Stoll and Louis, 2007; Stoll et al., 2006; Toole & Louis, 2002; Mitchell & Sackney, 2000). PLCs emphasize collaboration and reflection among a group of teachers to expose them to new ideas and practices, so they can improve on their pedagogy and their teaching through a process of critical inquiry. The concept of PLC originated and has been studied and more widely applied in the context of developed countries, mainly the United States, Netherlands and England. There is only limited research on PLCs in the context of developing countries (some exceptions include Lee & Lee, 2018; Lee et al., 2011; Zhang & Pang, 2016; Tatto, 1999; Pryor, 1998; Avalos, 1998; and Abrahams, 1997), which is mostly qualitative in nature (except for Lee & Lee, 2018; and Lee et al., 2011).

Although PLCs are in vogue in many high and low resource contexts, there is limited experimental and quasi-experimental evidence of their impact on teaching and student learning. To my knowledge, only one review conducted by Vescio, Ross and Adams (2008) assessed the effect of PLCs on teaching practices. The authors found that all 11 research articles reviewed, which were mostly qualitative in nature, supported the idea that well-developed PLCs can have a positive effect on teaching practices (Vescio, Ross & Adams, 2008). Overall, experimental (or quasi-experimental) and/or long-term designs are nearly absent in the literature (exceptions include Supovitz, 2002; Dunne et al., 2000) and there are only a few quantitative cross-sectional studies available (exceptions include Sargent, 2014; Lee et al., 2011) The few quantitative studies that have investigated the relationship between participation in PLCs and teaching outcomes have found mixed results. For instance, Sargent (2014) tests the relationship between frequency of participation in school-level PLCs and teachers' reported use of innovative pedagogy. She finds that teachers who participate more frequently in PLC activities in the school are more likely to report frequent use of innovative teaching methods. Supovitz (2002) compared team-based (a form of PLC) and non-team-based teachers on their instructional practices. Contrary to the findings of Sargent (2014), he does not find any difference between elementary school teachers in team-based and non-team-based schools on their individual instructional practices (Supovitz, 2002).

trained. The main reason given for the lack of effectiveness of the school-based training was the heavy workload of all teachers which left little time for systematic input.

Overall, there is a lack of quantitative studies on PLCs, especially in the context of low-and middle-income countries. There is also not enough research assessing the extent to which existing Western conceptualizations of PLCs are valid in low-resource settings. Further research is needed to improve our understanding on how PLCs work in developing countries and how different collaborative practices affect teaching and learning outcomes.

In-school pedagogical support by principals

Principals can play a key role in providing in-service professional development to teachers through school-based pedagogical support. By supporting teachers and influencing teaching quality, classroom conditions, and the effective use of resources within the school, principals can directly influence teaching and indirectly student achievement (Hallinger & Heck, 1999; Hallinger, 2005). Indeed, instructional leadership has been identified by the Western literature as a key characteristic of effective principals and as a strong predictor of student learning. In their meta-analysis of 27 studies, Robinson et al. (2008) conclude that instructional leadership actions are among the strongest predictors of student achievement, and that "*the closer educational leaders get to the core business of teaching and learning, the more likely they are to have a positive impact on students' outcomes*" (p. 664). Important aspects of instructional leadership highlighted in the literature involve building relationships with teaching staff and focusing such relationships on some very specific pedagogical work (Robinson et al., 2008). There is also evidence from the United States showing that providing training to principals in how to improve interaction with teachers —by providing feedback to teachers on lesson plans, action plans to improve student performance, and classroom behavior— can lead to a large impact on student learning (Fyer, 2017).

More broadly, it has been widely recognized that a good principal is the key to a successful school (Branch et al., 2012). Indeed, an emerging literature suggests that school principals are critical for ensuring academic achievement and in developing high-quality schools (Horng et al., 2010; Andrews & Soder, 1987). However, in contrast to the large literature on teacher quality (Araujo et al., 2016; Chetty et al., 2014; Rivkin, et al., 2005; Rockoff, 2004; Harris & Sass 2006; Kane et al., 2008; Buddin & Zamarro 2009), fewer studies have addressed the impact of principals in the production of student achievement. Emerging evidence from the United States suggest that highly effective principals have the potential to raise student achievement (Branch et al., 2012; Dhuey & Smith 2014; Coelli & Green 2012). Some studies have attempted to empirically identify specific skills that principals need to promote student achievement but found mixed evidence. Clark and colleagues (2009) find no relationship between principal education and students' standardized exam scores. On the other hand, Eberts and Stone (1988) and Ballou and Podgursky (1993) find that more educated principals are associated with worse school performance.

Most of the academic literature on the impact of principals and principal's instructional leadership on student achievement emerged in the context of developed countries, and it is unclear to what extend the findings are also applicable to developing countries. Given the uniqueness of the contexts, social norms and structural conditions found in developing countries, further research is warranted. As highlighted by Hallinger (1999), it is imperative to understand how different cultures reshape popularized concepts and practices of principals. Additionally, most of the available literature examined the effect of principals on composite measures of student performance, and only a few studies have tested the effect of principals by subject area (Dhuey & Smith, 2014; Clark et al., 2009).

1.4. Problem statement

The expansion of the quantity and availability of schools in low-and middle-income countries came at the cost of the quality of education, generating a learning crisis (Hanushek, 1995). To meet the increased enrollment needs, there has been a massive recruitment of less educated and/or untrained teachers, which has led to concerns about the quality of education being provided (Orr et al., 2013). In Sub-Saharan African specifically, the problem of low learning achievement emerges in the early grades, as the teaching of reading, which is crucial to children's progress through school, is highly ineffective (Bashir, 2018).

Because of the learning crisis, over the past decade international and national agencies have begun to emphasize the need to improve education quality rather than quantity in developing countries (Wagner et al., 2012). Improving teaching quality to ensure that all children in school actually obtain the skills and knowledge they are meant to acquire have become key objectives (UNESCO, 2010). Nonetheless, problems still persist in both pre-service and in-service teacher training. Weak teacher pre-service education systems result in teachers lacking subject knowledge and pedagogical skills when entering the teaching profession (World Bank, 2018). In-service teacher trainings in low- and middle-income countries, usually held in a central location, are also often not effective in improving teacher practice (Popova et al., 2018). If teachers in developing countries have the opportunity to participate in professional development, it is often a one-off training in a central location. This is often the case in Sub-Saharan Africa, where teacher training is delivered through short courses centrally planned by the Ministry of Education on specific topics, such as the introduction of new curricula (Lauwerier & Akkari, 2015), and are not connected to the context or the on-the-ground reality teachers experience in their school communities. Possible reasons for the failure of teacher training programs in developing countries may rely on their focus on imparting knowledge, on being overly theoretical, and on their delivery being too rote and passive (Loyalka, forthcoming). All these characteristics make it difficult for these training programs to be translated into practical pedagogical improvements as teaching is a skill that needs to be developed through ongoing practice (Kennedy, 2016).

In order to continuously improve teaching quality in Sub-Saharan Africa for in-service teachers and change pedagogical practices, the literature recognizes that it is critical for continuous support to be provided close to the school or within the school and be related to improving instruction (Bashir, 2018; Hardman et al., 2011; Hardman, 2009). Woods and McQuarrie (1999) argue that job-embedded learning (learning that occurs in the school as teachers and administrators engage in their daily work activities) is likely to be most effective as it allows for concrete problems faced in the local environment to be raised, and teachers can receive feedback on their actual teaching. Within school continuing professional development support for teachers can take multiple forms, including: coaches or mentors who offer constructive feedback; peer-to-peer interactions and collaboration through communities of practice at their school; and teacher interaction and support from their school head on pedagogical matters (Bashir, 2018). Despite the recognized importance of within school continuing in-service professional development, as highlighted in the previous section, still little is known on how collaborative efforts between teachers themselves and between teachers and school heads or coaches-mentors can influence teaching practices and student achievement.

In this dissertation, we contribute to the existing literature by investigating the relationship between different types of within and across school collaboration practices on teaching and student learning. From a general point of view, we focus on (i) collaboration within school between teachers on pedagogical topics; (ii) principals' instructional support to teachers; and (iii) teachers' collaboration through PLCs. We empirically investigate teachers' collaboration and principal support in Equatorial Guinea, while PLCs are explored in three Sub-Saharan African countries: Equatorial Guinea, Ghana and Nigeria. Furthermore, we develop a typology of PLCs that incorporates elements that are specific to how PLCs function in Sub-Saharan Africa. We discuss the objectives and specific research questions in more detail next.

1.5. Research objectives and research questions

As noted previously, the objective of this research is to examine the influence of collaborative practices within schools and between schools on student achievement and teaching practices. The main research questions and sub-questions this thesis addresses are highlighted in table 1.1. While the main questions guide the primary analysis in the empirical chapters, the sub-questions explore heterogeneity issues (e.g., urban versus rural effects, effects by ethnic groups) or mechanism through which the main question operates. By addressing both the main questions and sub-questions proposed, each of the empirical chapters fill some of the existing gaps in the literature. For the purpose of this thesis, collaboration is grouped into two categories: (i) collaboration on pedagogical topics at the school level arising organically among teachers themselves and between teachers and the principal, addressed in chapters four and five; and (ii) PLCs, as a more structured approach to

teachers' collaboration within a school or cluster of schools, addressed in chapters six and seven of this thesis.

The research questions proposed are relevant given that research and practitioners are praising and promoting school-based continuing teacher professional development as alternative or complementary forms to traditional professional development approaches. For example, Hardman et al. (2011) highlight a shift in Eastern and Southern Africa away from institutional-based primary teacher education towards more flexible school-based provision. According to the authors, the use of in-service trainings based in the school or in a cluster of schools "*has been strongly advocated as a way of closing the gap between theory and practice and raising the quality of teaching and learning in the region's primary schools*" (p. 670). As a result, governments in the region have been setting up national in-service education and training (INSET) strategies and continuous professional development (CPD) systems for teachers (Hardman et al., 2011).

Yet, very little is known about the influence of collaborative practices, which form the backbone of in-service professional development at the level of individual schools or cluster of schools, on teaching and student achievement, especially in the context of in the context of Sub-Saharan Africa. As highlighted at the beginning of this introduction, there is a lack of quantitative studies investigating the influence of teachers' collaboration with their peers or participation in more structured PLCs on teaching and student learning. The literature on principal's effect on student learning is more robust but is based mostly on the experiences of developed countries and it is unclear to what extend these findings are also applicable to developing countries. Given the deep education challenges and inefficiencies in education systems in Sub-Saharan Africa, it is critical to investigate if teachers' collaboration and principal instructional support to teachers can also influence teaching and student learning in these settings. The lack of literature on within school and across collaborative practices in developing countries generally, and in Sub-Saharan Africa specifically, motivates further research on this topic.

It is also pertinent to explore the influence of collaborative practices – both among teachers as well as between teachers and the principal in the Sub-Saharan African context - because we simply do not know if they can promote transformation changes in teaching practices or if they work to conserve existing traditional substandard practice. On one hand, existing research suggests that effects of teachers' collaboration and successful principal leadership may be greater in schools that are in more difficult circumstances (Leithwood et al., 2004). This would suggest a larger effect of teachers' collaboration and principals' instructional support on teaching and learning in low-resources countries. On the other hand, it is important to acknowledge that in Sub-Saharan Africa countries pedagogical practices may be resistant to change due to a discrepancy between what is required for teachers to teach under learner-centered approaches and the preparation they receive in

initial teacher education, which is highly teacher-centered (Altinyelken, 2011; Akyeampong et al., 2013). For instance, if principals provide instructional leadership to preserve and reinforce rooted and ineffective teaching practices, they may not have an indirect effect on student achievement.

Table 1.1. Main research questions

	Main question	***Sub-questions***	***Related chapter***
1	Does the Equatorial Guinea Primary Education Survey respond to existing best practices and quality standards in survey methodology?		Chapter 2
2	Do the reading and mathematics assessments used in the Equatorial Guinea Primary Education Survey respond to quality standards in the assessment literature?	Are the reading and mathematics assessments used in the Equatorial Guinea Primary Education Survey reliable?	Chapter 3
3	Does teachers' level of collaboration affect student achievement?	Does teachers' likelihood to collaborate differ based on teachers' ethnic background?	Chapter 4
4	Does principal's instructional support to teachers contribute to student achievement?	To what extent is the influence of principal's instructional support practices different for language and mathematics achievement? To what extent is the influence of principal's instructional support practices on student achievement different for urban versus rural areas?	Chapter 5
5	To what extent PLCs in developing countries differ from existing theoretical frameworks proposed by the Western literature?	What elements are typical of PLCs in developing countries? Can these elements be integrated into existing theoretical frameworks?	Chapter 6
6	Are higher levels of participation in PLCs associated with higher constructivist and student-centered teaching?		Chapter 7
7	Is PLC strength associated with higher constructivist and student-centered teaching?	Through what mechanisms does the strength of a PLC influence teaching?	Chapter 7

The questions in table 1.1 are studied using a mix of quantitative and qualitative methods. *Questions 1 and 2* will be addressed through a review of the survey methodology and student learning assessment literature as well as through psychometric analysis. *Questions 3 and 4* will be addressed econometrically using different techniques, including: instrumental variable, pseudo-panel, coarsened exact matching, cluster robust inference, and two-level intercept estimates. The application of these econometric techniques will be explained in detail in *chapters 4 and 5* of this dissertation. *Question 5* will be addressed using a multi-method data collection and analysis approach explained in detail in *chapter 6*, encompassing document analysis, semi-structured interviews and validation sessions with experts. Through this qualitative approach we are able to create a typology of PLCs integrated into a conceptual framework, which reflect modes of PLC operation that are specific to Sub-Saharan African countries. Having this theoretical backbone

allows us to empirically explore *questions 6 and 7* in *chapter 7*, through regression analysis with school fixed effects.

1.6. Research setting

The main research setting that is common across all empirical chapters in this thesis is Equatorial Guinea. For this reason, we devote more attention in the introduction to describing this setting. Equatorial Guinea is a small country in West-Africa, made up of a mainland territory called Rio Muni, and five islands including Bioko, where the capital Malabo is located. Data from 2015 Census points to a population of approximately 1.2 million people, although this number must be interpreted with caution, as it represents an increase of approximately 50 percent relative to previous estimates (Bassett et al., 2017). According to the World Factbook (2017) the country's population is divided as following: Fang 85.7%, Bubi 6.5%, Mdowe 3.6%, Annobon 1.6%, Bujeba 1.1%, other 1.4%. Official languages of the country include Spanish (adopted since 1844 and is the official language of education and administration), French (made official in order to join the Francophonie) and Portuguese (adopted in 2010 to gain membership in the Community of Portuguese Language Countries). Local languages spoken in the country include Fang, Bube, Ndowe, Bissio, and Pichinglis.

The country was a Spanish colony until 1968. Under Spanish colonization, the economy was dominated by agriculture (cocoa) and forest (wood) sectors (Bassett et al., 2017; McsSherry, 2006). Foreigners, mostly Spanish and Nigerian, controlled the small service sector, while the vast majority of Equato-Guineans remained subsistence farmers with little or no integration into the market economy (McsSherry, 2006; Wood, 2004). In the post-colonial period, both agriculture and forestry experienced a sharp decline (Bassett et al., 2017).

After gaining independence in 1968, a 11-year authoritarian regime took place which nearly destroyed the country's infrastructure and closed its schools (Lipski, 2004). Francisco Macias Nguema, the country's first president, came to power in 1968, installing one of post-colonial Africa's most oppressive regimes (McsSherry, 2006). During his presidency, Macias imposed a violent crackdown on the opposition, resulting in thousands of prisoners, murders, and 160,000 exiles (Bassett et al., 2017; Razquin, 2006; McsSherry, 2006). The period was marked by period marked by brutal dictatorship, high levels of social and political conflict, and significant human rights abuses (Bassett et al., 2017; Razquin, 2006). By the end of his regime, one third of the population had been killed or exiled (Wood, 2004). The economy also suffered: plantations controlled by foreigners were closed, Nigerian contract workers and Spanish expatriates expelled from the country, resulting in the cocoa industry virtually disintegrating (Bassett et al., 2017; Razquin, 2006; McsSherry, 2006; Wood, 2004).

Macias filled most government posts with members of his family and the Esangui clan of the Fang group, which originates from the mainland Rio Muni (McsSherry, 2006). Nguema discriminated and persecuted the Bubi people, the dominant ethnicity on Bioko island, increasing ethnic tensions between the Fang and Bubi (McsSherry, 2006). Tensions remain to this date, and there is deep ethnic division in Equatorial Guinea, and also between clan divisions within ethnicities (Genocide Watch, 2012). Nguema also began murdering members of his own clan and family, which eventually led to his nephew, Teodoro Obiang Nguema, undertaking a successful coup in 1979 and coming to power under the title "liberator" (McsSherry, 2006). Obiang's brother had recently been murdered at request of Macias, which may have triggered the coup (Wood, 2004).

Obiang remains president today and nepotism and kleptocracy among the president's family and the Esangui clan has led to a concentration of the nation's oil wealth and political power (Genocide Watch, 2012). Since oil-discovery in the mid-1990s the country has become one of sub-Sahara's biggest oil producers, but a large proportion of its population still lives in poverty. According to data from the International Monetary Fund, purchasing power parity Gross Domestic Product (GDP) per capita in Equatorial Guinea in 2016 was US$36,017[8], the highest in Africa and the 31st highest in the word. Nonetheless, in 2015 the country was ranked 138th out of 188 countries on the United Nations Human Development Index by the UNDP (2015).

Obiang' regime ended the reign of terror but maintained many of the repressive policies and corruption exiles (Bassett et al., 2017; McsSherry, 2006). Razquin (2006) characterizes the present situation as a 'restricted democracy', since despite the democratic structure, an opposition is nearly nonexistent. The government frequently detains the few opposition politicians that remain in the country, and cracks down on any civil society groups that are being politically engaged (Freedom House, 2017). Human Rights groups consistently rank the Equatorial Guinea at the bottom of political and civil freedom indexes. The country is ranked as one of the nine least free countries in the world by Genocide Watch (2012). Freedom House's 2017 annual global survey scored Equatorial Guinea as a 7 for both political rights and civil liberties., where scale 1 represents most free country and 7 least free country.

Education System

The first lasting phase of colonization, which began around 1883, can be characterized as Catholic. In 1887, the Spanish administration officially entrusted the Claretian mission (a community of Roman Catholic of priests and brothers) with everything related to teaching and education in Equatorial Guinea (Negrín Fajardo, 2011). Although there were also some public educational

[8] International Monetary Fund. Report for Selected Counties and Subjects. Retrieved from: https://www.imf.org/external/pubs/ft/weo/2018/01/weodata/weorept.aspx?sy=2016&ey=2023&scsm=1&ssd=1&sort=country&ds=.&br=1&pr1.x=47&pr1.y=4&c=642&s=NGDPRPC%2CNGDPRPPPPC%2CNGDPPC%2CNGDPDPC%2CPPPPC%2CGGR%2CGGX&grp=0&a=#cs5

establishments in the colony, they were functioning irregularly. There was a competition for the control of colonization between the different representatives of the Spanish state administration in Equatorial Guinea and the Claretian mission. The mission was strongly supported in Madrid by conservative and ecclesiastical groups, who would limit the governor's authority, when not respective to the privileges of the Claretian missions (Negrín Fajardo, 2011). The implementation of the Second Spanish Republic (1931-1939) was marked by a change in the colonial education policy, supported by new legislative rules and the application of fundamental principles of secularism and religious freedom. One of the most contentious issues was the subjugation of the Claretian mission, who lost some of the power and budget they had been receiving from the state (Negrín Fajardo, 2011). However, with the end of the Spanish civil war and the start of Francoist Dictatorship in 1939, the new regime started implementing in Spain an ideology of national-Catholicism, which was also extended to colony (Negrín Fajardo, 2011). Specifically, it declared obligatory the teaching of Catholic religion and sacred history and that the school should be presided over by the image of the Crucified One and the portrait of the Head of State (Negrín Fajardo, 2011).

During 1968–1979, because the education sector was considered an opponent of the ruling regime, the country experienced the dismantling of the physical and technical-organizational infrastructure of the education sector (PRODEGE, 2011). During this period, teachers and school administrators were targeted for repression (Razquin, 2006) and some teachers lost their lives and / or took the path of exile (PRODEGE, 2011). The schools were reopened shortly after a coup in 1979 that lead to the formation of a new government, but the effects of the 11-year regime are still visible (Krause, 2010).

Equatorial Guinea' education system is characterized by low student learning outcomes and inefficiencies – a large numbers of learners in the early primary grades lack foundational reading and mathematics skills, while primary level drop-out and repetition rates are high (World Bank, 2017). Half of primary students are overage for their grade, 24% of primary students repeat a grade, and only 22% of students eligible for secondary school ultimately enroll (UNICEF 2016). As illustrated in Table 1.1, gross and net enrollment ratios in primary education are low (61.58% and 43.34% respectively), indicating that a sizeable proportion of the primary school-age population is still out of school. For those who are in school, survival to the last grade of primary is low - only 72% of students who enroll in first grade are expected to reach the last grade of primary education - and repetition rates are high at 18.52%. Students in school are also not showing signs of learning – a large number of learners in the early primary grades still lack foundational reading and mathematics skills (Bassett et al., 2017).

Several factors contribute to the poor quality of education in Equatorial Guinea, including low teachers' qualification, teachers' absenteeism, lack of learning materials and poor physical conditions. There are low rates of qualified teachers in the system: 5 percent of teachers in preschool,

54 percent in primary school, and 69 percent in secondary school are qualified. The supply of qualified teachers may also be compromised, as between 2006 and 2015 there was a steep decline of the enrollment in higher education schools (Bassett et al., 2017). There are also no mechanisms in place to hire or promote teachers based on their performance. While no official statistics are available on teachers' absenteeism, it is a commonly cited problem, together with lack of teachers' interest in their work (Bassett et al., 2017). Workshops conducted by PRODEGE in 2016 identified that neither school principals nor academic department coordinators are supervising the issue of absenteeism (*Ibid*). Teachers must often cope with large groups of pupils of different ages all grouped together under a single class (UNESCO, 2004). Traditional classroom instructional practices tend to dominate, as teachers depend heavily on dictation and copying and make limited use of problem solving and dialogues initiated by the students (Bassett et al., 2017). The situation is even more acute in rural areas. A report from UNESCO (2004) highlights that schools in rural areas have virtually no resources; some do not even have furniture. In rural areas, often there is no electric power, no drinking water, no latrines. There are schools that do not have their own premises, and in some villages, pupils attend class in the chapels of one of the churches.

Table 1.2. Key Primary Education indicator in Equatorial Guinea in 2015, by gender

Indicator	**Total**	**Female**	**Male**
Gross enrollment ratio (%)	61.58%	61.34%,	61.81%
Net enrollment rate (%)	43.34%	43.82%,	42.88%
Percentage of repeaters in primary (%)	18.52%	19.57%	17.44%
Survival to the last grade of primary (%)	72.07%	72.16%	72.06%
Gross intake ratio into the last grade of primary (%)	40.6%	39.65%	41.57%

Source: UNESCO Institute for Statistics, 2015

Despite these challenges, education sector spending has remained low as government investments allocation are skewed towards infrastructure (roads, ports, airports) instead of productive and social investment. According to the World Bank (2010), recurrent education expenditures in 2008 accounted for only 11 percent of total current expenditure, representing 0.2 percent of GDP. This is far lower than expenditure in other sub-Saharan Africa countries that allocate 3.9 percent of GDP (World Bank, 2010). Between 2003 and 2008 public expenditure in education accounted for 7 percent of public expenditure, compared to 16 percent in the neighboring countries (Gabon and Cameroon) and 30 percent in Uganda and Tanzania, and 25 percent in Ghana (*Ibid*). Another issue relies in that State was only financing wages for 59 percent of the entire teaching staff (6,038 teachers), based on data from 2008 (*Ibid*). For teachers being financed by the State, the median wage

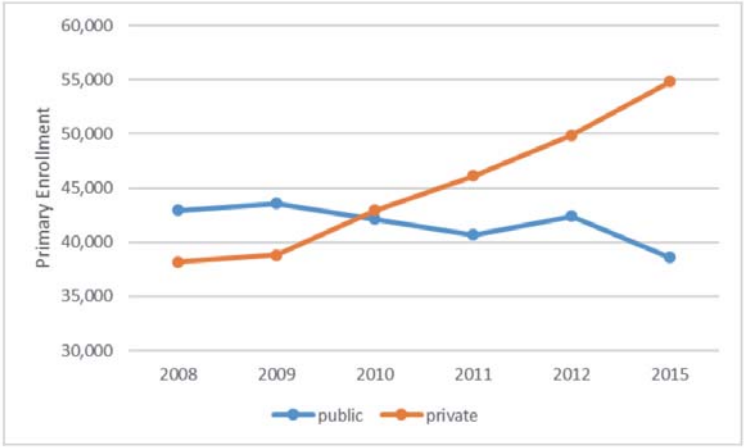
as a percentage of per capita GDP was 13 percent in 2008 (*Ibid*). As a comparison, in Sub-Saharan Africa the median wages are four times as large as per capita GDP. For example, average wages for teachers are seven times as large as per capita GDP in Burundi, four times larger in Cameroon, and five times larger in Côte d'Ivoire, Kenya, and Nigeria (*Ibid*). It is also worth noting the disparity in the wage system, as school managerial staff earns seven times more than teachers (*Ibid*). Overall, State shortfalls in education funding have been covered by other sources of financing, such as the payment of enrollment fees at private centers, contributions from Parent Teacher Associations at public centers (which often goes towards paying volunteer teachers), as well as international assistance from external donors, such as the Spanish Agency for International Cooperation for Development (AECID) and Amerada Hess Foundation, which finances a portion of PRODEGE (*Ibid*).

Spain's influence in Equatorial Guinea's education system and specifically the influence of Catholicism, which started during the colonial period with the Claretian mission, is reflected in Spain's current contributions to education expenditures in the country and its allocation pattern. AECID contributed €6.9 million in 2008 (representing approximately 31 percent of all expenditure on education in that year), targeted particularly for primary schooling (World Bank, 2010). A significant share of this sum (€4.1 million) was allocated to education centers managed by the Spanish Federation of Religious Schools (Federación de Escuelas Religiosas de España) and the Association of Catholic Education Centers of Equatorial Guinea (Asociación de Centros Católicos de Enseñanza de Guinea Ecuatorial) (*Ibid*).

In Equatorial Guinea, education centers can be private or public. While there are more public than private centers in total numbers, there are more students enrolled in private centers than in public ones (Bassett et al., 2017). There is also a trend observed in recent years characterized by a dramatic increase in primary education enrollment in private centers and a slight decrease in enrollment in public centers (Figure 1.5). However, it is important to clarify that there are very few "pure" private schools in Equatorial Guinea, that is, schools that do not receive any support from the government (about 8 percent of primary schools and 4 percent of secondary schools were classified as private receiving no public support in 2010-2011) (Bassett et al., 2017). Most private schools receive some form of government subsidy, such as assignment of teachers on the government payroll to their teaching staff. There are also several annual transfers from the public budget and investment projects to private centers, which are authorized by law (Bassett et al., 2017). Overall, private centers are more independent from the Ministry of Education and Science (MoES) in the sense that they have pedagogic autonomy to adopt and implement programs, methodologies, and management systems. However, they must comply with approval, supervision, and control from the MoES (*Ibid*). The growing share of private schooling could be seen as a way to increase supply (and possibly quality),

but could also be seen as a symptom of problems in the public-school sector (although private schools are not necessarily of better quality, even if they cost more for families) (*Ibid*).

Figure 1.5. Total Primary Enrollment in Equatorial Guinea by Sector, 2008-2015



Source: Bassett et al. (2017), using data from PRODEGE/MEC Education statistics Yearbooks (Anuarios Estadisticos), 2007-08 through 2014-15.

Education decision making is highly concentrated in the MoES offices in Malabo, on the island of Bioko. Because high level MEC central office positions are used for patronage, at the policy setting level (Directors General and above) the MEC central office has high turnover and low expertise. Overall, the MEC does not have effective norms, structure, or resources in place. The MEC central office is protocol driven because it operates using norms that are mostly tradition and process oriented, submitting most decision making to formal commissions and groups or waiting for approval from the Minister. There are no mechanisms in place to appoint, hire or promote principals and teachers based on their qualifications or on the job performance (Bassett et al., 2017). The MEC central office assigns principals and teachers to public schools in an arbitrary manner without a previous consultation process with the principals or teachers that will be dispatched. There is also no indication that there is a systematic process that guide the choice of which teacher or principal is assigned to each school. Principals have no mechanisms to hire, evaluate, and promote their teaching staff based on performance, which leaves a quality 'vacuum' and the lack of autonomy and accountability regarding teacher management (Bassett et al., 2017).

Equatorial Guinea as a research setting

There are a couple reasons why Equatorial Guinea is an especially interesting setting to study the influence of collaborative approaches on teaching practices and student achievement. First, as we mentioned, the country's education system is characterized by low teachers' qualifications, traditional pedagogical approaches and low student learning outcomes (Bassett et al., 2017).

Exploring alternative and sustainable methods to teacher professional development that can lead to improved teaching and learning is key for the country to improve the delivery of quality education. Second, for initial grades of elementary schools, on which we focus, we have reliable and detailed data on student achievement in math and language as well as on student's, teacher's and principal's background characteristics. Specially in developing countries, "data suitable for doing rigorous empirical work in this area are scarce" (Grissom & Loeb, 2011).

Although this dissertation focuses mostly on Equatorial Guinea, the challenges the country faces in primary education are similar to the challenges faced by other Sub-Saharan African countries (e.g., low student learning, low teachers' qualifications, teachers' use of traditional pedagogical approaches in the classroom, among others). In addition, given trends observed across the region towards long-term sustainable vision of continuous professional development based on flexible approaches grounded in the school or cluster of schools (Hardman et al., 2009), the findings from Equatorial Guinea are also relevant for other countries in the region. However, the findings from the dissertation should be interpreted with caution, as it is not possible to know the external validity of the results to other contexts.

1.7. Research outline

This dissertation is comprised by eight chapters. For didactic purposes we group the chapters under three main categories, represented by parts I, II and III. **Part I** focuses on data collection aspects and consists of two chapters on the Equatorial Guinea Primary Education Survey (EG-PES), which provides the data for the empirical analysis under chapters 4 and 5. The EG-PES was developed and implemented by the Program for National Educational Development (PRODEGE) of Equatorial Guinea, which is implemented by the Non-Government Organization Family Health International 360 (FHI360) in partnership with the Ministry of Education and Science (MoES) of Equatorial Guinea. As the Monitoring and Evaluation Specialist for PRODEGE, the PhD candidate was responsible for leading all technical and logistical aspects of the EG-PES, including sampling and instrument design, hiring and training of enumerators, coordination of field logistics, data processing, analysis and reporting. Since the quality of a survey is of prime importance for accurate, reliable and valid results, Chapter 2 analyzes how the EG-PES responds to quality standards in survey methodology, while chapter 3 addresses how EG-PES' measurement approach responds to quality standards in student learning assessments. **Part II** focuses on assessing the impact or influence of school-based peer collaboration and principal instructional support on student learning. It consists of two stand-alone empirical papers, which are presented in Chapters 4 and 5. **Part III** focuses specifically on PLCs as a form of collaboration within and across schools and shifts the focus towards teaching outcomes, and how participation in PLCs are associated with specific teaching practices. Part III also consists of two stand-alone papers, which are presented in Chapters

6 and 7. Overall, both **parts II and III**, which encompass the four empirical papers of this dissertation, investigate how collaborative practices and structures within schools or across schools are associated or impact teaching and student learning. By looking at different collaborative practices under each chapter, we improve our understanding of how specific approaches can potentially inform in-service continuous professional development policies and approaches.

Part I: Chapters 2 and 3

Chapter 2 reviews the main aspects emphasized by the survey methodology literature on how to improve the overall quality of a survey by minimizing survey errors. The aspects reviewed in this chapter include sampling design, questionnaire design, respondent and interviewer effects and processing of survey data. The specific case of EG-PES is presented and analyzed after each sub-section to determine how this survey responds to the quality standards highlighted in the literature.

Chapter 3 provides an overview of measurement approaches to student learning and socio-economic status. The chapter reviews student learning assessments in the developing world and standards for quality of effective assessments. The specific case of a hybrid learning assessment used in the EG-PES is explored in detail, with a description of its development, piloting and reliability and analysis of how it responds to the quality standards identified. Given the recognized role that family background and household socioeconomic status play in student learning, the chapter also introduces the computation of a wealth index using data on household characteristics and durable asset ownership collected by the EG-PES.

Part II: Chapters 4 and 5

Chapter 4 explores the influence of teachers' collaboration on student achievement in Equatorial Guinea using an instrumental variable approach. An instrumental variable approach is used in this chapter to deal adequately with issues of omitted variable bias and endogeneity. Specifically, we use exogenous variation in the number of teachers per school as an instrument for the endogenous variable teachers' collaboration. Using a 2SLS estimation with a latent-variable model applied in the first-stage of the 2SLS procedure, the instrumented variation in teachers' collaboration is used to carve out the impact on student achievement. In addition, we explore the heterogeneous effects across different population groups. Given the deep ethnic divisions in Equatorial Guinea, we explore if teachers' likelihood to collaborate differ based on teachers' ethnic background.

Chapter 5 investigates the influence of principals' instructional support to teachers on student achievement. Equatorial Guinea is a unique setting to investigate this relationship. The education system in the country is highly centralized and the Ministry of Education and Science (MoES) assigns principals to public and private schools arbitrarily without undergoing a consultation process. This process mimics a natural experiment, as principal allocation to schools and hence to

students is determined arbitrarily by an outside entity, resembling a random assignment. Due to the lack of longitudinal data in Equatorial Guinea, in this chapter we construct a pseudo-panel using a Coarsened Exact Matching approach. Specifically, we match third grade students with first grade students based on key observable characteristics. We estimate the effect of instructionally supportive principals on student achievement using an ordinary least squares regression model, controlling for principal, teacher and student background characteristics, including student predicted past achievement. We employ cluster-robust standard errors that corrects the standard errors for within-cluster error correlation as well as a two level-random intercept model, which accounts for students nested in classrooms and classrooms nested within schools.

Part III: Chapters 6 and 7

Chapter 6 aims to inductively create a typology of PLCs that incorporates elements that might be specific to how PLCs operate in low- and middle-income countries, with a focus on Sub-Saharan Africa. This study employs a multimethod approach, encompassing document analysis, semi-structured interviews with PLC experts and expert validation sessions. The hypothesis we aim to test with our qualitative approach is that PLCs in developing countries may operate in different ways, reflecting their unique context, society and culture, not currently captured in the conceptualizations of the Western Literature.

Chapter 7 builds of the typology and conceptual framework introduced in chapter 6 and uses data from a teacher survey implemented with over 2,300 preschool and primary education teachers in Equatorial Guinea, Ghana and Nigeria. First, the chapter examines the relationship between high levels of participation in PLCs and constructivist student-centered teaching using Ordinary Least Square (OLS) regression. This model includes school-fixed effects to account for unobserved differences between schools and controls for teacher background characteristics and number of months since the PLC has been established. Second, the chapter tests the relationship between the strength of a PLC using a PLC Strength Index and constructivist student-centered teaching practices also employing OLS regression and the same control variables as the previous model. We measure PLC strength using teacher's self-reporting of the presence of PLC core characteristics informed by the Western literature and incorporated in the conceptual framework introduced in chapter 6. Because the PLC Strength Index is a variable at the school level, we are unable to include school-fixed effects, but we employ cluster-robust standard errors. Third, we expand this model to include three PLC structural varying characteristics from our conceptual framework and explore how these features interact with PLC strength to influence constructivist student-centered teaching.

Conclusion

Chapter 8 presents the main conclusions of this dissertation, which are presented around statements. This allows the identification of promising collaborative practices and approaches that can influence teaching and student learning. The chapter also discusses the study limitations, as well as implications of the findings for future research and policy.

Table 1.3. Dissertation Outline

	Chapter	Research Questions	Methods	Country
	Chapter 1	General Introduction		
Part I: Overview of Data Collection	Chapter 2	Does the Equatorial Guinea Primary Education Survey respond to quality standards in survey methodology?	Review of the survey methodology literature	Equatorial Guinea
	Chapter 3	Do the reading and mathematics assessments used in the Equatorial Guinea Primary Education Survey respond to quality standards in the assessment literature?	Review of assessment literature; Reliability analysis; Exploratory factor analysis; Principal component analysis	Equatorial Guinea
Part II: School-based collaboration	Chapter 4	Does teachers' level of collaboration affect student achievement?	Instrumental Variables	Equatorial Guinea
	Chapter 5	Does principal's instructional support to teachers contribute to student achievement?	Pseudo-panel through Coarsened Exact matching; Cluster robust OLS estimate; Two level-random intercept estimate	Equatorial Guinea
Part III: Professional Learning Communities	Chapter 6	To what extent PLCs in developing countries differ from existing theoretical frameworks proposed by the Western literature?	Document Analysis; Semi-structured Interviews	Democratic Republic of Congo, Equatorial Guinea, Ghana, Nigeria, Senegal, South Africa
	Chapter 7	Are higher levels of participation in PLCs associated with higher constructivist and student-centered teaching? Is PLC strength associated with higher constructivist and student-centered teaching?	OLS with fixed-effects	Equatorial Guinea, Ghana & Nigeria
	Chapter 8	Conclusion		

References

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and Student Achievement in the Chicago Public High Schools. *Journal of Labor Economics*, *25*(1), 95–135.

Abrahams, M. (1997). Collaborative space in South African schools: A comparative perspective. *Journal of Negro Education, 66*(4), 409-422.

Akyeampong, K., Lussier, K., Pryor, J., & Westbrook, J. (2013). Improving teaching and learning of basic maths and reading in Africa: Does teacher preparation count? *International Journal of Educational Development*, *33*(3), 272–282.

Altinyelken, H. K. (2010). Pedagogical renewal in sub- Saharan Africa: the case of Uganda. *Comparative Education*, *46*(2), 151–171.

Andrews, R. L., & Soder, R. (1987). Principal leadership and student achievement. *Educational leadership*, *44*(6), 9-11.

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

Aslam, M., & Kingdon, G. (2011). What can teachers do to raise pupil achievement? *Economics of Education Review*, *30*(3), 559–574.

Avalos, B. (1998). School-based teacher development the experience of teacher professional groups in secondary schools in Chile. *Teaching and Teacher Education*, *14*(3), 257-271.

Ballou, D., & Podgursky, M. (1995). What makes a good principal? How teachers assess the performance of principals. *Economics of Education Review*, *14*(3), 243-252.

Bau, N., & Das, J. (2017). *The misallocation of pay and productivity in the public sector: Evidence from the labor market for teachers*. Washington, DC: The World Bank Group.

Bashir, S., Lockheed, M., Ninan, E., & Tan, J. P. (2018). *Facing forward: Schooling for learning in Africa*. Washington, DC: The World Bank Group.

Bassett, L. K., Di Gropello, E., Marshall, J. H., Abril Tabares, J. A. (2017). *Equatorial Guinea education sector diagnostic (English)*. Washington, D.C.: The World Bank Group.

Bold, T., Filmer, D., Martin, G., Molina, E., Stacy, B., Rockmore, C., ... & Wane, W. (2017). Enrollment without learning: Teacher effort, knowledge, and skill in primary schools in Africa. *Journal of Economic Perspectives*, *31*(4), 185-204.

Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2012). *Estimating the effect of leaders on public*

sector productivity: The case of school principals (No. w17803). National Bureau of Economic Research

Bruns, B. & Luque, J. (2014). *Great teachers: how to raise student learning in Latin America and the Caribbean*. Washington, DC: The World Bank Group.

Bruns, B., Costa, L., & Cunha, N. (2017). *Through the looking glass: can classroom observation and coaching improve teacher performance in Brazil?.* Washington, DC: The World Bank Group.

Buddin, R., & Zamarro, G. (2009). Teacher qualifications and student achievement in urban elementary schools. *Journal of Urban Economics*, *66*(2), 103-115.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014a). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, *104*(9), 2593-2632.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014b). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American economic review*, *104*(9), 2633-79.

Clark, D., Martorell, P., & Rockoff, J. (2009). *School Principals and School Performance* (Working Paper 38). National Center for Analysis of longitudinal data in Education research.

Coelli, M., & Green, D. A. (2012). Leadership effects: School principals and student outcomes. *Economics of Education Review*, *31*(1), 92-109.

Chesterfield, R., & Abreu-Combs, A. (2011). Centers for Excellence in Teacher Training (CETT): Two-Year Impact Study Report (2008-2009). *Washington, DC: Oficina de USAID para América Latina y el Caribe*.

Conn, K. M. (2014). *Identifying Effective Education Interventions in Sub-Saharan Africa: A meta-analysis of rigorous impact evaluations* (Doctoral dissertation, Teachers College).

Dhuey, E., & Smith, J. (2014). How important are school principals in the production of student achievement?. *Canadian Journal of Economics/Revue canadienne d'économique*, *47*(2), 634-663.

Eberts, R. W., & Stone, J. A. (1988). Student achievement in public schools: Do principals make a difference?. *Economics of Education Review*, *7*(3), 291-299.

Evans, D. K., & Popova, A. (2015). *What really works to improve learning in developing countries? An analysis of divergent findings in systematic reviews* (Policy Research Working Paper No. 7203). Washington, DC: The World Bank Group.

Freedom House. (2017). Freedom in the World 2017. Populists and autocrats: The dual threat to global democracy. *Freedom House*.

Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation*, *28*(2), 465-498.

Fuller, E. J. (1999). *Does teacher certification matter? A comparison of TAAS performance in 1997 between schools with low and high percentages of certified teachers*. Austin: Charles A. Dana Center, University of Texas at Austin.

Fryer Jr, R. G., & Levitt, S. D. (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics*, *2*(2), 210-40.

Genocide Watch (2012). Updated country profile of Equatorial Guinea. Retrieved June 30, 2019, from http://genocidewatch.net/2013/03/19/country-profile-equatorial-guinea/

Glewwe, P., & Kremer, M. (2006). Schools, Teachers, and Education Outcomes in Developing Countries. *Handbook of the Economics of Education 2*. 945- 1017.

Grissom, J. A., & Loeb, S. (2011). Triangulating principal effectiveness: How perspectives of parents, teachers, and assistant principals identify the central importance of managerial skills. *American Educational Research Journal*, *48*(5), 1091-1123.

Hallinger, P. (2005). Instructional leadership and the school principal: A passing fancy that refuses to fade away. *Leadership and policy in schools*, *4*(3), 221-239.

Hallinger, P., & Heck, R. (1999). Can leadership enhance school effectiveness. *Educational management: Redefining theory, policy and practice*, 178-190.

Hanushek, E. A. & Rivkin S. G. (2006). Teacher Quality. Developing Countries. *Handbook of the Economics of Education 2*. 1051-1075.

Harbison, R. W., & Hanushek, E. A. (1992). *Educational performance of the poor: lessons from rural Northeast Brazil*. Oxford University Press.

Hardman, F., AbdKadir, J., Agg, C., Migwi, J., Ndambuku, J., & Smith, F. (2009). Changing pedagogical practice in Kenyan primary schools: the impact of school- based training. *Comparative Education*, *45*(1), 65-86.

Hardman, F., Ackers, J., Abrishamian, N., & O'Sullivan, M. (2011). Developing a systemic approach to teacher education in sub-Saharan Africa: Emerging lessons from Kenya, Tanzania and Uganda. *Compare: A Journal of Comparative and International Education*, *41*(5), 669-683.

Harris, D., & Sass, T. (2006). The effects of teacher training on teacher value-added. *Florida State University. Unpublished.*

Horng, E. L., Klasik, D., & Loeb, S. (2010). Principal's time use and school effectiveness. *American Journal of Education*, *116*(4), 491-523.

Junaid, M. I., & Maka, F. (2015). *In-service teacher education in Sub-Saharan Africa: A synthesis report*. Commonwealth Secretariat. London. Ministry of Education (1996), Educating Our Future: National Policy on Education, Lusaka: Zambia.

Kremer, M., Brannen, C., & Glennerster, R. (2013). The challenge of education and learning in the developing world. *Science*, *340*(6130), 297-300.

Lauwerier, T., & Akkari, A. (2015). *Teachers and the quality of basic education in sub-Saharan Africa*. Education Research and Foresight: Working Papers Series, (11). UNESCO: Paris.

Lee, J. C. K., Zhang, Z., & Yin, H. (2011). A multilevel analysis of the impact of a professional learning community, faculty trust in colleagues and collective efficacy on teacher commitment to students. *Teaching and teacher education*, *27*(5), 820-830.

Lewin, K. M. (2009). Access to education in sub- Saharan Africa: Patterns, problems and possibilities. *Comparative Education*, *45*(2), 151-174.

Leithwood, K., Seashore, K., Anderson, S., & Wahlstrom, K. (2004). *Review of research: How leadership influences student learning*. Minneapolis: University of Minnesota, Center for Applied Research and Educational Improvement

Lucas, A. M., McEwan, P. J., Ngware, M., & Oketch, M. (2014). Improving early grade literacy in East Africa: Experimental evidence from Kenya and Uganda." *Journal of Policy Analysis and Management, 33*, 950–976.

Lomos, C., Hofman, R. H., & Bosker, R. J. (2011). Professional communities and student achievement–a meta-analysis. *School effectiveness and school improvement*, *22*(2), 121-148.

Loyalka, P., Popova, A., Li, G., & Shi, Z. (Forthcoming). Does Teacher Training Actually Work? Evidence from a Large-Scale Randomized Evaluation of a National Teacher Training Program. *American Economic Journal: Applied Economics*.

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education review*, *27*(6), 615-631.

Kennedy, M. M. (2016), 'How does professional development improve teaching?', *Review of*

Educational Research. 86(4), 945–980.

Krause, J. (2010). *The Business of State Building: The Impact of Corporate Social Responsibility on State Development in Equatorial Guinea*. Dissertation, University of Florida.

Lee, J. C. K., Zhang, Z., & Yin, H. (2011). A multilevel analysis of the impact of a professional learning community, faculty trust in colleagues and collective efficacy on teacher commitment to students. *Teaching and teacher education*, *27*(5), 820-830.

Lee, D. H. L., & Lee, W. O. (2018). Transformational change in instruction with professional learning communities? The influence of teacher cultural disposition in high power distance contexts. *Journal of Educational Change*, *19*(4), 463-488.

Lipski, J. (2004). The Spanish language of Equatorial Guinea. *Arizona Journal of Hispanic Cultural Studies*, *8*(1), 115-130.

Malpel, J. (2016). PASEC 2014: Education system Performance in Francophone Sub-Saharan Africa. *Programme d'Analyse des Systèmes Educatifs de la CONFEMEN. Dakar, Sénégal*.

McEwan, P. J. (2015). Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments. *Review of Educational Research*, *85*(3), 353-394.

McSherry, B. (2006). The Political Economy of Oil in Equatorial Guinea. *African Studies Quarterly*, *8*(3).

Mitchell, C., & Sackney, L. (2000). *Profound improvement: Building capacity for a learning community.* Lisse, The Netherlands: Swets & Zeitlinger.

Michaelowa, K. (2001). Primary education quality in francophone Sub-Saharan Africa: Determinants of learning achievement and efficiency considerations. *World development*, *29*(10), 1699-1716.

Negrín Fajardo, O. (2011). ¿ Qué queda en Guinea Ecuatorial de la educación española? de la educación colonial a la realidad actual. *Hist. educ*, 30, 111-126.

Orr, D., Westbrook, J., Pryor, J., Durrani, N., Sebba, J., Adu-Yeboah, C. (2013). *What are the impacts and cost-effectiveness of strategies to improve performance of untrained and under-trained teachers in the classroom in developing countries?* London: EPPI-Centre, Social Science Research Centre, Institute of Education, University of London.

Piper, B., Zuilkowski, S. S., & Mugenda, A. (2014). Improving reading outcomes in Kenya: First-year effects of the PRIMR Initiative. *International Journal of Educational Development*, *37*, 11-21.

Piper, B., & Korda, M. (2010). *Early Grade Reading Assessment (EGRA) Plus: Liberia. Program Evaluation Report*. Durham, NC: RTI International and USAID.

Piper, B. (2009). *Integrated Education Program: Impact Study of SMRS Using Early Grade Reading Assessment in Three Provinces in South Africa*. Produced for review by the United States Agency for International Development. Prepared by RTI International, North Carolina.

Popova, A., Evans, D. K., Breeding, M. E., & Arancibia, V. (2018). *Teacher Professional Development around the World: The Gap between Evidence and Practice* (Policy Research Working Paper, 8572). Washington, D.C.: The World Bank Group.

PRODEGE (Programa de Desarrollo Educativo de Guinea Ecuatorial). (2011). *Cambios en el Desempeño Docente en Educación Primaria*. Washington D.C.: FHI360.

Pryor, J. (1998). Action research in West African schools: problems and prospects. *International Journal of Educational Development*, *18*(3), 219-228.

Razquin, P. (2006). *Informe de Situación del Sector Educación y Lineamientos para Fortalacer la Política Educativa*. Washington D.C.: Banco Mundial.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417-458.

Robinson, V. M., Lloyd, C. A., & Rowe, K. J. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational administration quarterly*, *44*(5), 635-674.

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

Sailors, M., Hoffman, J. V., Pearson, P. D., Beretvas, S. N., & Matthee, B. (2010). "The Effects of First- and Second-Language Instruction in Rural South African Schools." *Bilingual Research Journal, 33*, 21–41.

Sargent, T. C. (2014). Professional learning communities and the diffusion of pedagogical innovation in the Chinese education system. *Comparative Education Review*, *59*(1), 102-132.

Snilstveit, B., Stevenson, J., Phillips, D., Vojtkova, M., Gallagher, E., Schmidt, T., ... & Eyers, J. (2015). *Interventions for improving learning outcomes and access to education in low-and middle-income countries: a systematic review*. London: International Initiative for Impact Evaluation.

Strauss, R. P., & Sawyer, E. A. (1986). Some new evidence on teacher and student competencies.

Economics of Education Review, *5*(1), 41-48.

Stoll, L., Bolam, R., McMahon, A., Wallace, M., & Thomas, S. (2006). Professional learning communities: A review of the literature. *Journal of educational change*, *7*(4), 221-258.

Stoll, L., & Louis, K. S. (2007). *Professional learning communities: Divergence, depth and dilemmas*. Berkshire, UK: Open University Press

Supovitz, J. A. (2002). Developing communities of instructional practice. *Teachers College Record*, *104*(8), 1591-1626.

Tatto, M. T. (1999). Improving teacher education in rural Mexico: The challenges and tensions of constructivist reform. *Teaching and teacher education*, *15*(1), 15-35.

Tolley, L. M., Johnson, L., & Koszalka, T. A. (2012). An intervention study of instructional methods and student engagement in large classes in Thailand. *International Journal of Educational Research*, *53*, 381-393.

Toole, J., & Louis, K. S. (2002). The role of professional learning communities in international education. In K. Leithwood & P. Hallinger (Eds.), *Second international handbook of educational leadership and administration* (pp. 245–281). Dordrecht, The Netherlands: Kluwer.

United Nations Development Program (2015). *Human Development Report 2015. Work for Human Development*. New York, NY: Author.

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2010). *Reaching the marginalized* (EFA Global Monitoring Report 2010). Paris, France: Author.

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2013). *The Global Learning Crisis: Why Every Child Deserves a Quality Education*. Paris, France: Author.

UNESCO Institute for Statistics (UIS). (2018). Handbook for measuring equity in education. Montreal, Canada: Author.

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2004). *Multigrade Schools: Improving Access in Rural Africa?* Paris, France: Author.

United Nations Children's Fund (UNICEF). (2015). *Global Out-of-School Children Initiative Operational Manual*. New York: United Nations Children's Fund (UNICEF), Education Section, Programme Division.

UNESCO Institute for Statistics (UIS). (2016). *Leaving No One Behind: How Far on the Way to Universal Primary and Secondary Education?*. (Policy Paper 27/Fact Sheet 37) Montreal,

Canada: Author.

UNESCO Institute for Statistics (UIS). (2006). *Teachers and Educational Quality: Monitoring Global Needs for 2015*. Montreal, Canada: Author.

Woods, F., & McQuarrie, F. (1999). On the job learning. *Journal of Staff Development, 20*(3), 10-13.

Wood, G. (2004). Business and politics in a criminal state: the case of Equatorial Guinea. *African Affairs*, *103*(413), 547-567.

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

*The World Factbook (*2017). Washington, DC: Central Intelligence Agency, 2017.

Wagner, D. A., Lockheed, M., Mullis, I., Martin, M. O., Kanjee, A., Gove, A., & Dowd, A. J. (2012). The debate on learning assessments in developing countries. *Compare: A Journal of Comparative and International Education*, *42*(3), 509-545.

World Bank (2010). *Equatorial Guinea: public expenditure review (English)*. Public expenditure review (PER). Washington, DC: World Bank.

Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and teacher education*, *24*(1), 80-91.

Chapter 2. The Equatorial Guinea Primary Education Survey: A Quality Assessment

2.1. Introduction

The quality of a survey is of prime importance for accurate, reliable and valid results (Ustun et al., 2005). Different aspects of surveys, such as the design, collection, processing and analysis, can affect the quality of the survey estimates ((De Leeuw et al., 2008). Chapters four and five of this dissertation rely on one specific survey source, the Equatorial Guinea Primary Education Survey (EG-PES). The EG-PES was developed and implemented by the Program for National Educational Development (PRODEGE) of Equatorial Guinea, which is implemented by the Non-Government Organization Family Health International 360 (FHI360) in partnership with the Ministry of Education and Science (MoES) of Equatorial Guinea. As the Monitoring and Evaluation Specialist for PRODEGE, the PhD candidate was responsible for leading all technical and logistical aspects of the EG-PES, including sampling and instrument design, hiring and training of enumerators, coordination of field logistics, data processing, analysis and reporting. The main objective of the EG-PES is to collect information on teachers' classroom instructional practices and students' competences on reading and mathematics. The survey is comprised of a battery of instruments: 1) student's background questionnaire; 2) reading and mathematics learning assessment; 3) teacher's background questionnaire; 4) teacher's classroom observation; 5) principal interview questionnaires.

Chapters four and five rely in the EG-PES data to respond the following research questions:

1. To what extent does teachers' level of collaboration affect student achievement?
2. To what extent does the principal's instructional support to teachers contribute to student achievement?

Our main dependent variable (student achievement) as well as the independent variables of interest (teachers' level of collaboration and instructional support to teachers) are computed using data collected by the EG-PES.

Given the reliance on this survey as the main source of data to respond to both research questions, in the present chapter we analyze key sampling and administration procedures adopted by the EG-PES in light of the survey methodology literature. Our goal is to assess if the EG-PES responds to existing best practices and quality standards and if it can be a trusted data source for this research.

Much of the survey methodology literature revolves around the question of why errors arise in survey and how they can be minimized. According to the definition proposed by Groves (2004), survey errors refer to "deviations of obtained survey results from those that are true reflections of

the population" (Groves, 2004, p.6). Fowler & Mangione (1990) propose a similar definition: "error is the difference between a survey answer and the true value of what the researcher wants to measure" (Fowler & Mangione, 1990, p.24). By adopting a total survey error framework, survey methodologists recognize that each aspect of a survey has the potential to affect its results (Groves et al., 2011). The survey methodology literature divides sources of survey errors into two categories: errors of non-observation and observation or measurement errors. Errors of non-observation occur when the characteristics of the respondents do not match those of the population of interest, from which they were drawn. Measurement errors occur when the answers to the questions are not good measures of the intended construct (Groves et al., 2011).

This chapter reviews the main aspects emphasized by the survey methodology literature on how to improve the overall quality of a survey by minimizing survey errors. The aspects reviewed in this chapter include sampling design, questionnaire design, respondent and interviewer effects and processing of survey data. The specific case EG-PES is presented and analyzed after each sub-section to determine how this survey responds to the quality standards highlighted in the literature.

2.2. Theoretical aspects of sampling design and errors of non-observation

Sampling design and implementation is a critical element of the survey process. As Groves et al. (2011) explain it, a survey can have all its phases well designed and implemented, such as a well-developed questionnaire, highly trained and motivated interviewers, excellent field supervision and management, etc. However, if there are issues with the sampling, the researcher will not be able to make inferences to the population that was the target of the investigation. Sampling errors might occur, which refer to the "*possibility that members of a sample differ from the larger population because of chance variation in the composition of the sample*" (Blair et al., 2013, p. 89). A sampling error may arise because of issues in the sampling frame, in the sample selection, or even after the sample has been successfully selected. The total error for estimating a quantity of interest from a survey, is the sum of four components: total survey error = coverage error + sampling error + nonresponse error + measurement error (Lorh 2008).

Coverage errors may be due to undercoverage or overcoverage. Undercoverage errors may occur when units of the target population have no chance of being selected for the survey (de Leeuw et al., 2008; Lepkowski, 2005). Undercoverage may be due to the fact that not all elements of the target population are covered in the sampling frame[9] and as a consequence these eligible members of the population not covered in the sampling frame cannot appear in any sample drawn for the survey. This is the case with hidden populations, such as unregistered migrants, refugee populations, which

[9] As per Groves et al. (2011) definition, target population is "the group of elements for which the survey investigator wants to make inferences by using the sample statistics" (Groves et al., 2011, p. 69. The sampling frame is the set of materials, lists or procedures intended to identify the elements of the target population (Groves et al., 2011).

are not present in a sampling frame. Overcoverage occurs when a unit from the target population appears more than once in the sampling frame (this may occur, for instance, when a sampling frame results from the combination of several lists) or when individuals not in the target population are included in the sampling frame and are not screened out of the sample (Lorh, 2008; de Leeuw et al., 2008). Since the survey can only estimate the mean of the target population covered in the sampling frame, coverage bias may emerge if there is systematic difference between covered and undercovered on the items being measured (Blair et al., 2013; Groves et al., 2011).

Thus, a central concern for survey research is to identify how well the sampling frame covers the target population. Undercoverage is a difficult problem to identify. However, once it is known that there are population elements in your ideal survey population that do not appear in the sampling frame, one alternative is to consider additional frames to identify them and update the original frame (Groves et al., 2011).

Sampling error occurs because a sample is taken instead of surveying the entire population (de Leeuw et al., 2008; Lorh, 2008). In general, the amount of sampling error is a direct function of the number of units included the final sample (de Leeuw et al., 2008). Sampling error can be reduced by drawing samples that are large enough to produce the precision wanted (de Leeuw et al., 2008). The selection method for drawing a sample is also a key consideration. If the sample is not a probability sample (*i.e.,* does not use a random selection procedure), then statistical inference is not appropriate (de Leeuw et al., 2008). Additionally, if some population groups have disproportionately high or low chances of being selected given the sample selection method of choice, selection bias will occur. In these cases, a simple random sample is impractical and alternative methods to draw samples such that each individual or sampling unit in the population has an equal chance of appearing in the sample is desirable (Cochran et al., 1954). For instance, to minimize issues of selection bias, a probability sample design with stratification features can be used to assure representation of population sub-groups in the sample (Groves et al., 2004).

After sampling frames have been constructed, and elements have been drawn from the frame, non-response error might occur. Non-response error is due to unsuccessful attempts to obtain data for all sampled units on all questions (De Leeuw et al., 2008; Lepkowski, 2005). There are two types of nonresponse in surveys: unit nonresponse and item nonresponse (De Leeuw et al., 2008). Unit non-response might occur when selected elements refuse to participate in the survey or cannot be found or when the interviewer does not speak the respondent's language. If non-response is substantially different across different groups of the surveyed population (*i.e.,* not random), there may be systematic differences between respondents and non-respondents on the items being measured (Blair et al., 2013). Item-nonresponse refers to the failure to obtain information for one or more questions in a survey (de Leeuw et al., 2008).

The final type of error is measurement error, which occurs refers to the difference between a measured quantity and its true value. Measurement error is the focus of Chapter 5 of this dissertation.

Overall, a carefully design sample selection should have three basic features: a list of elements in the population (sampling frame); random selection of elements from the list, which will assure that all elements in the frame have the same probability of being selected; and some mechanism, such as stratification, to ensure that key subgroups of the population are represented in the sample (Groves et al. 2011; de Leeuw et al., 2008). To avoid issues with the sampling it is important to define the population of interest before drawing the sample, to have a sampling frame that covers the elements of the target population, to select a sample that fairly represents the entire population and to obtain data from as much of the selected sample as possible (Blair et al., 2013).

2.3. Sampling design in the Equatorial Guinea Primary Education Survey (EG-PES)

The objective of the Equatorial Guinea Primary Education Survey (EG-PES), conducted in May 2015, was to collect information on students' competences on language and mathematics. Given the large size of the population of interest, the target population of the survey was restricted to students and teachers in first, third and sixth grades of primary education in the country. These grades were chosen as they represent the first grade, the mid-grade and last of the primary education cycle in the country. The National Program for Educational Development was responsible for implementing the EG-PES. The program, which is public-private partnership between the Ministry of Educational and Science (MoES) and an international oil company, is implemented by an American NGO (Family Health International 360) in cooperation with the MoES. As the Monitoring and Evaluation Specialist for the NGO, the PhD candidate was responsible for leading all technical and logistical aspects of the survey. Specifically, I lead all technical efforts related to sampling design, development and piloting of different instruments for data collection, data processing, data analysis and dissemination of the results. In addition, I was responsible for developing a nation-wide data collection logistics plan, hiring and training enumerators, as well as for oversight and monitoring field data collection activities.

In determining the sample size, the EG-PES took into account the local capacity to implement the fieldwork for data collection, as well as relied on similar studies conducted on comparable countries. Based on existing early reading data for countries in Latin America, Southeast Asia and Sub-Saharan-Africa, RTI International (2009) recommends a sample of 475 students in 48 schools for a confidence level of 90%. The EG-PES followed RTI International (2009) recommendation on the sample size of students, but it increased the number of schools in the sample to increase variability between students on key learning outcomes (given that students in the same school tend to be more homogenous). EG-PES aimed to sample 60 schools in total: 30 schools that offer first and third

grades (first cycle of primary education) and 30 schools that offer first, third and sixth grades (first and second cycle of primary education). Within each school, the EG-PES objective was to collect data on 8 students per grade, totaling 480 grade 1 students, 480 grade 3 students and 240 grade 6 students, selected randomly. Since the researchers were mostly interested in collecting information about early grade reading, preference was given to have a larger sample for first and third grades.

The EG-PES adopted a sample design that complies with three basic features proposed by the survey methodology literature.

1. A list of elements in the population (sampling frame)

The most updated Primary School Census available at the time of the survey was used as a sampling frame. Since the School Census does not list all the student population, schools were used as clusters of frame elements. The Primary School Census in Equatorial Guinea started in 2007 and was conducted on an annual basis until 2011. The School Census contains the most up to date record of schools in the country. The Census started with a list of schools that were registered within the MoES and it was updated annually to include the schools that were created or that were closed. Although undercoverage is a difficult issue to identify, it is fair to assume that the School Census' database covers accurately the population of schools, as it contains most, if not all, existing schools in the country. However, given that the EG-PES was conducted over three years after the last School Census, new schools may have been established over that time period that are not included in the sampling frame. Following Groves et al. (2011) suggestion, an attempt was made to obtain additional school records from the MoES that could potentially be used to update the sampling frame. Unfortunately, such records were never provided. This means the final sampling frame may underrepresent newer, and likely smaller schools, established in fast growing areas, such as the capital city (Malabo) and the city of Bata.

2. A random selection of elements from the frame

As explained in more detailed under point 3 below, the researchers randomly selected the schools (primary sampling units) as well as the students (secondary sampling units) to ensure a probabilistic sample.

3. Some mechanism to ensure that key subgroups of the population are represented in the sample.

The EG-PES adopted a two-stage stratified random sampling by cluster strategy to select students and teachers from the target population. According to this strategy, the sample is probabilistic (randomly selected), stratified by cluster, and selected in stages. The primary sampling unit is the school (cluster) and the secondary sampling unit is the student or teacher. To ensure that schools are proportionately represented in the sample, the frame was stratified by the following variables: grade

cycle offered; province; status of province capital for the two major provinces. Given the higher level of development and urbanization in the two main provinces (Bioko and Litoral) in comparison with the remaining provinces, stratifying at the province level was key to ensure representation. Schools in the least developed provinces usually have less access to good infrastructure and less qualified teachers. Although we aimed to ensure ethnic groups representation in the sample, there was no information available on the schools' ethnolinguistic compositions to use in the stratification process. Additionally, ethnic groups in EG are not necessary geographically concentrated, and most schools have a combination of groups. To facilitate the fieldwork logistics, schools that only offer 1st or 3rd or 6th grades were excluded from the target population. The stratification method generated 14 strata. Table 2.1 details the number of schools in each strata.

Table 2.1. Sampling Frame: number of school observations in each strata generated through the stratification process

Cycle offered	**Province**	**Capital Status**
Schools offering first cycle only (first and third grades, but no 6th grade) (n=457)	Bioko Norte and Sur (n=50)	Capital (n=32)
		Remaining (n=18)
	Litoral (n=114)	Capital (n=60)
		Remaining (n=54)
	Centro Sur (n=83)	
	Wele Nzas (n=101)	
	Kie-Ntem (n=109)	
Schools offering first and second cycle (first, third, and sixth grades) (n=217)	Bioko Norte and Sur (n=81)	Capital (n=60)
		Remaining (n=21)
	Litoral (n=84)	Capital (n=75)
		Remaining (n=9)
	Centro Sur (n=17)	
	Wele Nzas (n=15)	
	Kie-Ntem (n=20)	

Source: Equatorial Guinea 2011-2012 Primary School Census

A two-stage stratified random sampling by cluster strategy was adopted by the EG-PES as opposed to a stratified random sampling approach for two reasons: 1) at the time of the survey, there was no enrollment data available at the student level (i.e., a sampling framing listing all students in the targeted grades); and 2) even if data was available, sampling at the student level would have implied a complex and cumbersome field logistics for data collection with a high number of schools to visit. One of the advantages of the cluster approach is that it reduces the number of locations to visit when in-person data collection is needed.

This two-stage random sample technique is employed by different standardized international learning assessments, including TIMSS and PIRLS. Both TIMSS and PIRLS draw a sample of schools at the first stage and one or more intact classes of students are selected from each of the sampled schools as a second stage (as opposed to sampling students across classes in the same grade and school) (LaRoch et al., 2016). However, while TIMSS – PIRLS assess all students in selected classes, the EG-PES only assesses 8 randomly selected students. This is most related to the one-one nature of the EG-PES learning assessment, which we discuss in more detail in Chapter 3, while TIMSS and PIRLS can be administered to the entire group of students.

The first stage of sample selection drew 66 schools: 33 schools that offered first cycle education and 33 schools that offered first and second cycle education. By increasing the initial sample size with 10% (from 60 to 66 schools), we aimed to avoid having to rely on replacement of schools in the event that there was non-response during the field data collection, which would affect the selection probabilities. School non-response during field data collection could be related to factors such as:

- A school no longer offers the selected grades to be surveyed
- School authorities do not allow for random sampling of subjects that participate in the study or want to influence their selection
- A school that refuses to participate in the study
- A school is closed during the survey period for unforeseen reason (flooding, infrastructure problems, etc.)

The number of schools to be selected in each strata had to be proportional to its student population size, but with a minimum of two schools per strata to ensure representation of that strata in the sample. Table 2.2 presents the size of each strata, the proportion of each strata to the total student population per cycle, as well as allocation of schools across strata. The allocation of schools across strata was not strictly proportional for several reasons, being: 1) rounding of the allocation; 2) accommodation of the minimum of two schools per strata; 3) prioritizing to reduce the number of schools from the capital, allowing us to add to the outside of the capital stratum of the same province instead of other provinces.

Within each strata, school size, measured by number of enrollments in 1st, 3rd or 6th grade, was used for probability proportional to size (PPS) sampling. This method helps balancing the probability of selection as it gives a greater probability of selection to larger units than to smaller units according to size measure. Hence, with the PPS approach, larger schools have higher probability of being sampled (Biemer & Christ, 2008). Given that the same number of students are sampled per school, each student in the population has about the same probability of being sampled. Thus, the higher probability of selection for larger schools is offset by the lower probability of selecting children within the large schools.

Table 2.2. Allocation of Schools per Strata

Cycle	Province	# schools	Total # students	Proportion	Allocation of schools	Allocation (rounded)
1	Bioko	18	503	0.026526738	0.88	2
1	Bioko Capital	32	3139	0.16554161	5.46	5
1	Litoral	54	1437	0.075783145	2.50	3
1	Litoral	60	5188	0.273599831	9.03	8
1	Centro Sur	83	2615	0.137907394	4.55	4
1	WeleNzas	101	2689	0.141809936	4.68	5
1	KieNtem	109	3391	0.178831347	5.90	6
1 Sub-Total			18962	1	33.00	33
2	Bioko	21	2778	0.040028242	1.32	2
2	Bioko Capital	60	22906	0.330052881	10.89	10
2	Litoral	9	1434	0.020662526	0.68	2
2	Litoral	75	28300	0.407775104	13.46	12
2	Centro Sur	17	3568	0.051411363	1.70	2
2	WeleNzas	15	5016	0.072275616	2.39	2
2	KieNtem	20	5399	0.077794268	2.57	3
2 Sub-Total			69401	1	33.00	33
Total		674	88363			66

Source: Author's own calculations.

The PPS sampling approach was chosen because we aimed to establish a fixed number of students to be sampled in each school. By establishing a fixed number of students per school, it was possible to determine a fixed amount of time each data collection group would spend in each school, which simplified field logistics. In addition, a fixed number was established to minimize sample selection errors in the field. An alternative sampling approach would have been to randomly select the schools per strata, subsequently determine their size, and based on their size determine the number of students to be proportionately selected within each school. However, having different number of students to be sampled in each school could have more easily generated sample selection errors in the field. To simplify field logistics and to avoid sampling errors, a number of 8 students per grade was determined and a PPS approach adopted.

In the second stage of sample selection, which happened at the school level at the time of the survey, the 8 students were randomly selected in each grade within each school. If the school only had one classroom of the selected grade, all 8 students were sampled from that classroom. In schools with two classrooms of the selected grade, 4 students were sampled from each class. If, however, the school had more than two classrooms of the same grade, considering morning and afternoon shifts,

two classrooms were selected randomly, and 4 students were randomly selected in each classroom. Teachers of the selected grades and classrooms were interviewed and observed.

Within each classroom, students were randomly selected by the field staff based on systematic sampling using a random start and a sampling interval based on the size of the classroom. This random sampling was applied using the order of rows or students' distribution in the classroom. Refusals were recorded to assess non-response and students were replaced randomly if necessary. Also, field staff collected information of absenteeism to assess whether this could bias the sample towards the better attending students. A Kish selection table was developed as an instrument to assist the field staff in selecting the classrooms and the students. Appendix 2.A presents a template of the Kish tables used during the fieldwork.

In principle, two-stage stratified random sampling by cluster strategy was adopted by the EG-PES, where schools are sampled with probability proportional to school size and classes and students are sampled randomly, provides student samples with equal selection probabilities. However, in practice, differential patterns of nonresponse can result in varying selection probabilities, requiring a unique sampling weight for the students in each participating class in the study.

In summary, as discussed in this section, the EG-PES has three basic features for sample design proposed by the survey methodology literature. Although some undercoverage might occur given the time gap between the last school census and the survey administration, it is reasonable to assume that the sampling frame covered most existing schools at the time of the survey. The EG-PES sample design also purposefully attempted to minimize any potential issue of selection bias by adopting a stratification procedure. Finally, schools, teachers and students were randomly selected through a rigorous two-stage technique, ensuring they have equal selection probabilities and that they accurately represent the national target population of schools, teachers and students.

2.4. Implementation of the sampling design in the field

Before the start of the data collection fieldwork, the survey team called all 66 schools that were randomly selected to verify if they remained open. In case no telephone number for the school was provided, the team called the local education inspector responsible for the school zone. This procedure was aimed to make the data collection work more efficient by avoid traveling to schools that were closed. Out of the 66 schools, three were closed.

All 63 schools that were opened accepted to participate in the survey during the fieldwork. However, even though they were opened, some of them did not offer one or more grades that were part of the survey. The fieldwork team also found some schools with not enough students enrolled to reach the eight students per grade target. These two factors contributed to the final sample being smaller than anticipated. In addition to these two factors, in some schools the interviewers ran out of time and

were not able to apply the reading and mathematics learning assessments to all eight students sampled. This was especially true for grade three learning assessment, which proved to be longer than originally calculated. Not being able to apply the assessment to all sampled students within a given school lead to lower response numbers than the original sample size of students. Overall, the EG-PES reached 63 schools, 481 grade 1 students, 467 grade 3 students, 247 grade 6 students and 199 teachers from grades 1, 3 and 6.

Non-response was almost inexistent in the survey. All 63 school directors agreed to participate in the research; only one teacher and two sixth grade students refused to participate. The positive response and absence of refusals at the school level is likely due to three factors. First, all data collection groups carried a credential, signed by the Minister of Education, authorizing them to conduct their work as well asking collaboration from the overall population. As it will be discussed in more detail in chapter 5, Equatorial Guinea has a highly centralized and hierarchically organized state. Most teachers and school directors in the country are hired by the Government, even those working in private schools. Even though the credential did not make it mandatory for schools to participate (and the right of refusal was made clear by the interviewer), it is believed that the credential positively influenced participation. Second, before the survey team started any work within the schools at a given municipality or village, they made sure to inform all local political and education authorities about the survey. The support from the local authorities also facilitated the positive response at the local level. Third, the National Program for Educational Development was responsible for implementing the EG-PES. The Program is well regarded by the educational community in the country, as it constantly distributes didactic materials to schools and conducts in-service teacher trainings. The good imagine of the program also contributed to interviewers being welcomed in the schools.[10]

Given the authoritarian and repressive nature of the political regime, one could think that fear could have been the main motivating factor to participate in the study. To explore this issue in more depth and to better determine if fear of rejecting to participate played a role in the low non-response rate, I conducted a discussion with three Guinean local staff who participated in the fieldwork and were directly in contact with school directors and participants. I first asked them why they thought the survey was well received by the schools and none of them rejected to participate. The three of them agreed that it was because of PRODEGE's good imagine and standing within the education community. They explained that the program is constantly distributing didactic materials for all schools in the country, offering trainings, and paying financial incentives to participants. They mentioned cases that happened in the field during the survey implementation where local

[10] Since the survey involved the participation of human subjects, both adults and minors, the EG-PES team sought approval from a Protection of Human Subjects Committee (PHSC). The PHSC granted approval to conduct the research in April 2015

governments or school directors asked interviewers for financial resources or support to the school (textbooks, infra-structure, etc.). However, they did not condition their participation in the survey to the receipt of these benefits. In such cases, the survey team informed the respondents that they would pass along the concerns/requests to the responsible authorities within the MoES and Program for Educational Development, but that they personally were not in a position to address the concerns/requests directly. I questioned the three local staff if fear of rejecting to participate could have influenced the decision of school directors and teachers to participate. The three staff answered that it was highly unlikely. They also emphasized that most of the times the school directors were very receptive, welcoming and genuinely happy to receive the survey team, especially in the remote areas. This is a key consideration, because if participation in the survey was driven by fear, the veracity and quality of the responses would be questionable.

Sampling at the school level performed well during the fieldwork. Some school directors and teachers tried to influence the random sampling by suggesting a few students to take the assessment, most likely the ones with highest achievement. However, the survey team adhered to the random sampling procedure, using a Kish selection grid method (Kish, 1949). The Kish method, which was originally design for randomly selecting members within a household to be interviewed using a pre-assigned table of random numbers, was adapted for selection of students within classroom (see Appendix 2.A). We did notice however a frequent error, that occurred with most supervisors' log. The team supervisors were responsible for collecting information on the number of enrollments per classroom, per grade. However, in schools with more than two classrooms, supervisors only collected information on number of enrollments for the two classrooms selected to participate in the survey. Since total enrollment per grade is a key variable to calculate the sampling weights, where this information was missing it was replaced with the number of enrollments reported in the 2014-2015 School Census. We not expect this to influence any outcomes, as it only impacts weights marginally.

2.5. Theoretical standards for questionnaire design and respondent effects

Most surveys use questionnaires as instrument, which can be defined as a standardized set of questions administered to respondents (Groves et al., 2004). Questionnaires, however, may influence the respondent's answer through several channels: wording of the questions, questions length, the difficulty of the words used, open versus closed questions, the position of the question in the questionnaire, among others (Bradburn & Sudman, 2011). Improving questionnaire design to reduce response error is considered one of the most cost-effective steps to improve the quality of survey data (Fowler, 1995).

The literature recognizes certain standards that good questions need to follow. First, the question must gather the information needed to address the research objectives (Groves et al., 2011). Second,

all people answering the question should understand it in a consistent way and in a way that is consistent with what the researcher expect it to mean (Fowler, 1995). If respondents differ from each other or from the researcher in their understating of the question, measurement errors will arise. Additionally, respondents should have the necessary knowledge to answer the question and be willing to give correct and valid answers (Fowler, 1995). Overall, questions that are easily and consistently understood and produce few cognitive problems to the respondent introduce less measurement errors than questions that are hard to understand or difficult to answer (Groves et al., 2011).

When questionnaires do not follow these recognized standards, respondents' errors may arise. Blair et al. (2013) list three main types of response errors: comprehension error, knowledge error and reporting error. Comprehension errors occur when respondents do not understand a question correctly or misinterpret it or if different respondents understand the question in different ways. Blair et al. (2013) point out that comprehension errors can be avoided by careful question design as well as pretesting to ensure the respondents understand the question as intended. In addition, conducting focus groups to learn how potential survey respondents think and talk about the survey topic may help improve questionnaire design (Groves et al., 2011).

Knowledge errors occur if respondents do not know the answer to the question, or cannot recall it accurately (Blair et al., 2013). Respondents have more difficulty recalling an activity or event when there is a long time period between the event and the time of the survey (Kasprzyk, 2005). It is tautological to say that one main strategy to minimize this type of error involves matching the time frame of the question to a plausible recall period for the phenomenon in question (Blair et al., 2013). Other strategies proposed in the literature include the use of memory aids (calendars, maps, etc.) to reduce recall bias and the use of bounding techniques (use of a significant date or event as a reference period) to reduce the effects of telescoping[11] (Kasprzyk, 2005).

Reporting errors occur when respondents do not provide an accurate answer to a question: they skip the question or provide a false response. Reporting errors are usually associated with sensitive questions, for which respondents may overreport socially desirable behaviors or underreport socially undesirable behaviors (Blair et al., 2013). The use of forgiving wording has been recognized as a strategy to improve the reporting of sensitive information by the respondent (Groves et al., 2011). The most effective approach to improve reporting, however, is the removal of the interviewer from the question and answer process, through self-administered questionnaires (Groves et al., 2011).

[11] Telescoping errors occur when respondents erroneously report an event within the reference period when it actually occurred outside of that period (Kasprzyk, 2005). This phenomenon reflects a perception that past events seem closer to the present than they actually are (Groves et al., 2011).

To avoid respondents' errors, a critical part of the survey design is to evaluate how the questions respond to the standards for designing good questions. According to Fowler (1995), there are three main methods to evaluate survey questions: focus group discussions, cognitive interviews and field pretests. Focus group discussions can be useful to evaluate what members of the target population know about the issues covered in the questionnaires, how do they think about these issues and what terminology do they use in talking about them (Groves et al., 2004). Focus groups can help the researcher ensure that the words used in the questionnaire are familiar and understood by the respondents.

Cognitive interviewing is a type of in-depth interviewing technic which is used to explore the mental processes respondents use to answer survey questions (Campanelli, 2008). Cognitive interviews cover a range a procedures, but usually ask respondents to think aloud while they answer questions or just after it (Campanelli, 2008; Groves et al., 2004). Probes are also commonly used as a cognitive interview technic and involve asking follow-up questions designed to reveal the response strategy (Groves et al., 2004).

Pretests are small scales pilots of the data collection that aim to evaluate the survey instrument and data collection protocols. Two techniques are commonly used in pretests: respondent debriefing and interviewer debriefing. In the respondent debriefing, the interviewer asks the respondent about the interview process and the clarity of the questions at the end of the interview. In the interviewer debriefing, interviewers detect problems with questions, response categories and skip patterns and report it to the investigator (Groves et al., 2004). In addition to these two technics, the data collected during the pretest areoften use to identify items with high rates of missing data, out of range values, inconsistencies between questions and items with little variance (Groves et al., 2004).

Each of the evaluative strategies described in this section are recognized by the literature as important steps in improving the quality of the survey and consequently the quality of the data. The strategies can also be used to provide valuable evidence about the quality of the measurements that result from the survey.

2.6. Questionnaire design in the in the Equatorial Guinea Primary Education Survey (EG-PES)

The EG-PES is comprised of a battery of instruments: a Stallings classroom observation instrument; reading and math learning assessments; a student interview questionnaire; a teacher interview questionnaire; and a principal interview questionnaire. While this chapter reviews the development and evaluation of the different interview questionnaires, chapter four of this dissertation will focus on the development and testing of education measurement instruments, such as the learning assessment tests. Since this dissertation did not rely on data from the Stallings classroom observation instrument, the design and administration of this tool is not described or analyzed.

EG-PES interview questionnaires were carefully designed by a team of experts to respond to the standards recognized by the literature for good question design (Blair. et al., 2013; Groves et al., 2004; Fowler, 1995). The EG-PES team wrote clear and simple questions, using vocabulary easily understood by the target population, in an attempt to avoid comprehension errors. The final questionnaires also included plausible recall periods that did not require the respondents to remember activities or events that occurred long before the survey. In general, the questionnaires limited recall periods to one year, although most recall questions were designed within the frame of one week. By limiting the recall periods, as recommended by Blair et al. (2013), the EG-PES team aimed to minimize knowledge errors.

Since a government program was responsible for conducting the survey, researchers were concerned that respondents would try to match their answers to what the government considered to be desirable. EG-PES interviewers were instructed to disclaim, before the start of the interview, that the information provided by the respondent would be strictly confidential and protected by codes. Interviewers also made clear to respondents that he or she would not receive any personal benefit as a result of the survey. Given that information to respond to the research questions posed by chapters four and five is self-reported – teachers and principals responding on how often they meet to collaborate and work together on pedagogical topics - these disclaimers by the interviewers are important in avoiding errors associated with socially desirable or undesirable responses, as pointed out by Blair et al. (2013). By instructing interviewers to position themselves as neutral and to properly inform respondents on the nature and objective of the survey, the EG-PES research team aimed to improve the quality of the responses and reduce bias. Additionally, there is no reason to believe that teachers and principals had incentives or unobserved motives to over report their collaboration practices, since the country does not provide any incentive to teachers who collaborate more frequently.

Once the EG-PES team designed the questionnaires, two of the three strategies proposed by Fowler (1995) were applied to evaluate how the questions perform against the standards for designing good questions. First, focus group sections were conducted with a team of eight primary education teachers to ensure that the terminology used in the questionnaires was familiar and understood by the different category of respondents: students, teachers and principals. The focus group sections were also useful for the research team to understand how the target population thinks about the issues covered in the questionnaire. One clear difference between the perception of the researcher and that of the target population refers to the concept of time. The original questionnaire included a question asking students how many minutes it takes them to travel from their homes to the school, as a proxy for distance. During the focus group discussions, participants pointed out that children would unlikely be able estimate travel time in minutes, as this was an unusual measure in the local culture. Thus, the question as originally proposed in the questionnaire would likely bear unreliable

findings. As a result of four focus group sections, all three questionnaires were adjusted according to the recommendations from participants.

Second, the student interview questionnaire was piloted on a sample of ninety-nine first, third and sixth grade students within six different schools. Because of a matter of limited time and resources, the teacher and principal interview questionnaires were not pretested. The EG-PES team used the interviewer debriefing technic, described by Groves et al. (2011), to detect issues with the questionnaire and made adjustments accordingly. Each of the four interviewers that applied the questionnaire during the pretest was responsible for writing a report summarizing any potential issue encountered during the pretest, such as questions not easily understood by respondents, terminology used that was not appropriate for the context, lack of response categories for given questions, inconsistent skip patterns, among others.

As an example, one issue that appeared in the pretest reports refers to the question of whether or not the child had seen his/her father or mother reading at home in the previous week. This is considered an indicator for home literacy environment. Many children in Equatorial Guinea do not live with the father, mother or both. As the question was phrased originally, the negative answer was not capturing the fact that the father or mother did not read in the previous week, but the fact that the father or mother did not livein the same house as the child. Questions were adapted to reflect the issues encountered during the pilot and detailed in the interviewers' reports. However, even after adjusting the questionnaire based on the pilot results, some issues were detected after the fieldwork was completed and the data were processed. Using the same example as above, over 60% of the children in the sample reported they saw the father or mother read in the previous week. This seems extremely high for the context. Children may have interpreted "reading" as the act of reading any note or sign, not necessarily actual reading materials such as books and magazines. Thus, the question still did not properly capture the intended construct (home literacy environment) and as such could not be included in the analysis.

In summary, questionnaire design in the EG-PES followed most standards and practices recommended in the literature. Asides from careful consideration given to questions' design, two evaluative strategies, focus groups and pretests, were undertaken to evaluate and improve the quality of the questionnaires. Even though best practices for questionnaire design were followed, the analysis of the final data using descriptive methods showed that some questions did not perform as well as expected. In those cases, the data collected through these questions were excluded from the analysis.

2.7. Interviewer effect

The literature recognizes that interviewers play a critical role in the data collection process and that his/their performance can influence the quality of the data collected (Alcser & Clemens, 2010;

Groves et al., 2011; Fowler & Mangione, 1990; Kasprzyk, 2005). Because of individual differences, interviewer characteristics and performance might affect the respondents' answers, resulting in interviewer-related error. The interviewer may introduce errors in survey responses by not reading the items as worded, by probing inappropriately[12], by skipping a question that should not be skipped, by asking a question that should be skipped, by omitting part of a question, by adding new information to a question, or by adding an inflection or tone of voice that conveys a point of view (Blair et al., 2013). To the extent that the enumerator can affect an answer, there will be a difference between the survey answer and the true value of what the researcher wants to measure. Thus, differences across enumerators and differences in enumerator behavior can introduce bias to the data (Berazneva, 2014).

The literature indicates that interviewer-related errors can be reduced by standardized interviewing; selection, hiring and training of interviewers; and control of interviewers' behavior through supervision. Standardize interviewing identifies ways to reduce the effects interviewers might have on the answers they obtain. According to Fowler & Mangione (1990), "the goal of standardization is that each respondent be exposed to the same question experience, and that the recording of answers be the same, too, so that any differences in the answers can be correctly interpreted as reflecting differences between respondents rather than differences in the process that produced the answer" (Fowler & Mangione, 1990, p. 14). Common standardized interviewing practices described in the literature include the following: reading questions as worded; probing nondirective; recording answers exactly as given; and refraining from expressing views or onions.

In addition to standardization practices, one can argue that the quality of an interviewer-administered survey depends, to a large extent, on the quality of the interviewers themselves (Alcser & Clemens, 2010). Although there is no available empirical evidence on the impact of interviewer recruitment and selection criteria on the quality of the data (Groves et al., 2011), most authors consider it is important to attract and retain the most qualified interviewers possible. One recommended practice is to recruit more than the number of interviewers needed for data collection to allow for attrition and the dismissal of candidates who prove to be unsuitable (Alcser & Clemens, 2010).

The recruited interviewers should undergo a training of sufficient length to cover interview skills and techniques to help standardize the interview process (Fowler, 1991; Ustun et al., 2005). Experimental evidence available shows that the length of the training received can have an impact on interviewers' performance (Groves et. al, 2011). Trainings should cover standardized question-asking, questionnaire format and conventions, probing and recording of answers (Alcser & Clemens, 2010). Training methods should include role-playing and at least one opportunity to conduct a real-

[12] Probes are the words used by the interviewer to obtain an adequate answer when the respondent fails to answer a question completely and adequately (Groves et al., 2011). Probing properly means to probe in a way that does not affect which answer is chosen by the respondent.

life interview before the start of the data collection (Ustun et al., 2005). Training can also be useful to sensitize interviewers regarding the importance of the integrity of the data collection protocols to reduce falsification or fabrication of data (Groves et. al, 2011). In summary, interviewer training can ensure a uniform application of the survey materials (Ustun et al., 2005) and increase the accuracy of the information collected.

Supervising interviewers' behavior during the data collection is considered one of the main strategies for reducing interviewer-related error in surveys (Berazneva, 2014). Random field visits by project leaders, announced or unannounced, help evaluate the interviewers' performance and adherence to data collection protocols and training guidelines. In addition, on site supervision may help reduce another form of interviewer error, cheating, which happens when the enumerator fails administer the questionnaire, or parts of the questionnaire, and fabricates the data (Blair et al., 2013). Groves et. al. (2011) present evidence from two studies[13] showing that data quality improved when interviewers were audio recorded. These results indicate that monitoring interviewer performance can positively affect the quality of the data.

2.8. Interviewing protocol in the Equatorial Guinea Primary Education Survey (EG-PES)

The EGPED aimed to recruit, hire and train a total 24 data collectors and 8 team supervisors, who were be responsible for: (1) applying learning assessment instruments and interview questionnaires to selected students; (2) conducting classroom observations and applying interview questionnaires to selected teachers; and (3) applying interview questionnaires to directors in the sampled schools. With the objective of minimizing interviewer related error, the EGPED followed the three main strategies proposed in the literature.

2.8.1. Interviewers recruitment

To recruit and retain the most qualified interviewers possible, the EGPED applied a three-stage recruitment process to hire data collectors. The first stage consisted in a call for applicants. The position was advertised in universities and professional associations and required a university degree. In the second stage, the EGPED team selected a group of 60 candidates based on an analysis of the curriculums received, and invited them to take a reading and mathematics test. The test the candidates had to take was very similar to the sixth-grade learning assessments applied as part of the survey. To ensure a balanced composition of interviewers by local language spoken, candidates were asked to list the languages they were fluent in. The initial idea was to match students and

[13] Billiet, J., and Loosveldt, G. (1988). Interviewer Training and Quality of Responses. Public Opinion Quarterly, 52, pp. 190-211.
Fowler, F. J., Mangione, T.W. (1990). *Standardized survey interviewing; minimizing interviewer-related error*. Sage Publications: Newbury Park, CA.

interviewers based on the student's native language, especially in cases where the student did not express himself well in Spanish.

The 32 candidates with the best performance on the test were selected to participate in the third stage, the interviewer training. Following Alcser and Clemens (2010) recommendation, a higher number of interviewers than needed were recruited to participate in the training to allow for dismissal of candidates whose performance was considered unsatisfactory. At the end of the training, twenty-two interviewers with the best performance were hired. The final team was composed of relatively young individuals (mostly under the age of 30) and all had at least some university education. The educational backgrounds were diverse and included teaching, engineering, business administration, accounting, economics and environmental sciences. For most of them it was the first time working in a data collection activity. Six out of the twenty-two interviewers were recruited from the island to work there and the remaining sixteen were from the mainland. Most interviewers who had completed their bachelor's degree demonstrated a better understanding of the instruments and the protocols taught during training.

A few issues came up during the recruitment process. First, the performance of most candidates on the reading and mathematics test was below satisfactory levels: only a few were able to score above 70%. This raised questions about the ability of the candidates to apply the learning assessment, especially to advanced grades in primary education. Second, the vast majority of the applicants only spoke Fang and Spanish (over 95% of all candidates). Only a few applicants reported speaking French, Bulbi, or Kombe. Another issue refers to the performance during the training. Unfortunately, only 23 candidates demonstrated having acquired the skills in standardized interviewing taught during the training, even though the survey team aimed to hire 24 interviewers. Because of that, the survey was not able to maintain a capable pool of alternates. Among the 23 that started, one dropped out of the survey in the first week, due to health issues. Due to these issues, the survey team only hired 22 data collectors, instead of the 24 originally planned. The 22 that were hired worked throughout the entire survey, and no absence was registered. The team of supervisors and the EG-PES permanent local personnel were instructed to also perform the work of the data collectors missing in order to achieve the sample size originally planned. Unfortunately, that meant these supervisors spent less time observing and providing feedback to interviewers. To partially compensate for that, survey coordinates were instructed to visit these survey groups more often, as feasible.

Out of the 22 interviewers that were hired, only one spoke Bulbi and one spoke Kombe. During the fieldwork, interviewers were allowed to speak in their native language with the children, although the assessment itself had to be applied in Spanish. Given the low number of Bulbi and Kombe speakers, it was not possible to match interviewers and students based on their native language, as

originally envisioned. This could have generated some interviewer related error, as personal attributes of the interviewer can influence the respondent's answers (Groves et. al., 2011). It possible that a child who spoke Fang felt more comfortable with an interviewer whose native language was also Fang than a child who spoke Bulbi or Kombe. This could have had a negative impact on the results for non-Fang speaking children.

The recruitment for team supervisors followed a different strategy. Given the key role played by the supervisors, preference was given to candidates who had significant previous experience working with surveys. The EG-PES team searched the National Program for Educational Development database for curriculums of individuals who participated as interviewers or supervisors in education data collections conducted in previous years. In addition, applicants to the post of data collectors with significant data collection experience were consider for the team supervisor post. Following these two channels, a total of 12 candidates were contacted and invited to participate in an interview with the EG-PES team. During the interview, candidates were evaluated based on knowledge of survey practices and protocols; ability to solve problems in the field; and leadership skills. Eight candidates were selected for the post following the interview.

2.8.2 Training of interviewers and standardized interviewing

Selected interviewers participated on a two-week training workshop. The first day of training offered a general introduction to the survey, the objectives of the study, filed work organization and a review of roles and responsibilities. By introducing the participants to the objective and importance of the study, the EG-PES team aimed to sensitize them to the importance of the integrity of the data to reduce falsification or fabrication of questionnaires, as recommended by Groves et al. (2011).

Following the first day, the interviewers were split in two groups and participated in two parallel trainings: one that focused on the instruments aimed at collecting data from students and the other on the instruments that collect data from the teachers. The supervisors participated in both sections as well as were offered two extra days of training, which focused on selection protocols, fieldwork organization and field logistics.

Training for interviewers on the classroom observation instrument (Stallings) and teacher interview questionnaire was spread over a period of seven days and included: (1) two days of training on the interview questionnaire and data collection protocols; (2) a day of training on the theoretical aspects of the observation instrument and scoring; (3) three days of practical application in the local schools; (4) a final day of teach-back process and a test to measure the level of learning and inter-rater reliability among the interviewers.

Training for interviewers on the learning assessment instrument an student interview questionnaire lasted nine days and included: (1) two days of training on the interview questionnaire and data

collection protocols; (2) three days of training on the underlying principles of the learning assessments, procedures for its administration and scoring; (3) three days of practical application in the local schools and groups discussions following that; (4) one final day of test to measure the level of learning and to ensure inter-rater reliability.

Both trainings emphasized the data collection protocols, which included guidelines to interviewers on the steps to follow when arriving at the school, on steps to follow on participant selection and replacement, on how to request consent from participants, and on how to maintain logs that aim to register, code, and control the number of individuals (teachers and students) that were interviewed per school. The EG-PES team trained participants on the main aspects highlighted in the literature (Alcser & Clemens, 2010): standardized question-asking, questionnaire format and conventions, probing and recording of answers. The EG-PES training team emphasized the importance to read the questions as worded throughout the training. During the training, interviewers practiced reading the questionnaires aloud to one another several times so that they become comfortable with reading the questions aloud and with the structure of the questionnaire, including its skipping patterns. Role playing activities were undertaken, in which trainees practiced by interviewing other trainees, as recommended by Ustun et al. (2005). As recommended by the literature (Ustun et al., 2005), trainees traveled to local schools and performed interviews, applied learning assessment tests and classroom observations for three days as part of their practice.

At first, participants showed difficulty in reading the items as worded and tried to ask questions in their own way. Some participants also showed some initial difficulty in understanding and following the skip patterns established in questionnaires. However, with constant reinforcement from the training team and practice, they were able to overcome some of these difficulties. Some words in the questionnaire were also changed as a result of the training and the feedback received from the participants.

During the practices at the local schools, the EG-PES training team noticed that some enumerators tended to help out the students in solving the assessment exercises. This was especially true for enumerators who had a background in teaching. Upon noticing this behavior, the training team in charge of the supervision immediately offered individual feedback to the interviewer and supervised him/her more closely. In addition, given that the practice sections were followed by group discussions and feedback, the EG-PES training team emphasized the importance of reading and applying the sub-tasks as worded and not offering any type of extra support to students. Since this problem was diagnosed at an early stage (at the training phase as opposed to during the data collection), it was possible for the EG-PES training team to adopt measures to prevent it and pay closer attention to it during the fieldwork.

2.8.3. Control of interviewers' behavior through supervision.

Given that interviewers' supervision during the data collection is considered one of the main strategies for reducing interviewer-related error in surveys (Berazneva, 2014), the EG-PES team included three layers of supervision. The data collection field personnel were organized in eight teams of four people each. Each team was be comprised by one supervisor, one interviewer responsible for teachers' data collection and two interviewers responsible for students' data collection. One of the main responsibilities of the team supervisor was to oversee the work of the enumerators to ensure they followed the data collection protocols and guidelines thought during the training process. In addition to the supervision conducted by the supervisors, four EG-PES permanent local personnel were assigned to supervise the work of two data collection groups each. The four permanent personnel were regularly in the field and were instructed to report back to the EG-PES coordination team if they verified that an interviewer was not following the established protocols. As a third layer of supervision, EG-PES coordination team made regular surprise visits to the data collection teams to verify compliance with field procedures, attachment to the sample of selected schools, correct selection of students and teachers, and correct application of survey instruments.

By constantly supervising the work of the interviewers, the EG-PES team aimed to prevent data fabrication or falsification, as suggested by Blair et al. (2013), and aimed to improve the overall performance of the interviewers and quality of the data collected.

2.9. Processing of survey data

Data process or data capture refer to the process of entering numeric data into electronic files (Groves et al., 2011). The type of data processing used is dependent on the mode of data collection. In computer-assisted surveys, the interviewer or the respondent enter the data directly into an electronic file. With paper-based questionnaires, data entry clerks conduct the data entry by keying the digits of the questionnaire one by one (Groves et al., 2011).

For paper questionnaires, errors may occur during the data processing if the data entry clerk enters the data incorrectly into the electronic files (Blair et al., 2013). Measures have been identified in the literature to avoid or minimize such errors. The most common strategy refers to double data entry to identify errors and inconsistencies (Blair et al., 2013; Ustun et al., 2005). There is available evidence showing that when this strategy is applied, the rate of entry errors is low (Groves et al., 2011). Another commonly applied measure is the use of data entry programs that provide quality check features, such as range checks and checks to ensure logical consistency of related codes (Ustun et al., 2005). When using computer-assisted surveys, quality check features can be included in the electronic questionnaires and become part of the data collection itself. Finally, it is recommended that the complete interview forms are revised and checked by a supervisor before the

start of the data entry, to verify that the answers are correctly recorded and to check for inter-item consistency (Blair et al., 2013; Ustun et al., 2005). When possible, respondents should be re-contacted to resolve apparent errors or inconsistencies (Blair et al., 2005).

2.10. Mode of data collection and data processing in the Equatorial Guinea Primary Education Survey (EG-PES)

The EG-PES adopted a dual mode of data collection. For the student learning assessments and interview, paper-based questionnaires were used. For the classroom observations and teacher interview, electronic questionnaires in Android Tablets were adopted. The choice of data collection mode was guided by the availability of financial resources and an analysis of which instruments were more suitable for a computer-assisted mode.

The EG-PES team used Qualtrics software to convert the classroom observation grid and the teacher interview questionnaire into an electronic format compatible with the Qualtrics application for Android. The electronic questionnaires included range checks and enforced skip patterns and required responses established in the original paper questionnaire. The data collected through the individual tablets were weekly uploaded to a master excel database and stored in a cloud system. The main advantage of using computer-based systems is that it avoids many of the data processing errors highlighted by Blair et al. (2013) and Ustun et al. (2005).

Qualtrics was also the software used to develop the computer program for data entry for the paper questionnaires. As recommended by the literature (Ustun et al., 2005), the data entry program included quality check features and mechanisms to reduce human error, known as forms. Forms were created in the program to only allow a certain data type to be entered into a field, such as specific ranges and consistencies checks. Forms also constrained specific fields through the use of of pull-down menus, ensuring that the data were entered in the correct locations. Since the keying stations had internet access, the data entered was automatically transmitted to a central database, which was password protected.

To ensure data quality, the EG-PES coordination team monitored all procedures for the data processing in order to avoid errors. Although double data entry was not adopted due to its cost, the EG-PES team conducted random checks of the data. On average, EG-PES team revised 3% to 5% of the questionnaires entered by each of the six key enters hired. In cases where discrepancies where found, the team reviewed the data entered with the key enter using the physical questionnaires to identify the source and reason for the discrepancy and avoid it in other entries.

As advised by Blair et al. (2013) and Ustun et al. (2005), before the data entry process started, team supervisors also conducted quality control procedures. During the data collection process, team supervisors were responsible for conducting a daily review of the paper questionnaires, revising if

the answers were correctly recorded, if there was no missing answer, if the skipping patterns were properly followed and if there was inter-item consistency. Where they found inconsistencies, they advised the interviewer to return to the interviewee to correct for the mistake.

2.11. Ethical Considerations

Since the survey involved the participation of human subjects, both adults and minors, the EG-PES team sought approval from a Protection of Human Subjects Committee (PHSC). A proposal was submitted to the PHSC, detailing the research design and objectives as well as the measures that would be taken to protect human subjects. The PHSC granted approval to conduct the research in April 2015. This section details the steps that were taken by the EG-PES to protect the human subjects participating in the research. The EG-PES requested oral consent from each person being interviewed, assessed or observed as part of the study. Appendix 2.B provides the oral consent scripts used by the interviewer to address each type of participant: principals, teachers and students. Due to the distance students travel to school and the work schedule of parents, the EG-PES team considered it unrealistic to obtain parental consent for each student. Moreover, since the students were randomly selected at the time of the survey, it was not possible to predict which students would require parental consent in advance – and it was not feasible to contact the parents once the interviewers arrived at the school. Because of that, the EG-PES team requested a waiver to the PHSC from obtaining parental consent, which was granted.

Overall, the interviewers were instructed to explain the purpose of the study to participants in a way that was understandable to them; answer any potential question; explain that there are no major potential risks or benefits associated with the study; clarify that the decision to participate was voluntary and that participants had the freedom to discontinue participation at any time. The interviewers assured to participants that their privacy and confidentiality was protected through the use of codes in the databases. If necessary, the information was given to the participants' mother-tongue language. The team supervisors were instructed to supervise the interviewers during the consent and assessment process to ensure that no coercion occurred.

To protect the confidentiality of the students, no identifying information was collected. Teachers' names were collected but were maintained in a separate, confidential database, with access restricted to EG-PES coordination staff only. Teachers' names were removed from data files that are used for analysis purposes and were stored in a separate and secure location only accessible to the main investigators.

2.12. Concluding Remarks

Throughout this chapter, we provided a review of the main aspects emphasized by literature on how to improve the overall quality of a survey by minimizing survey errors. A discussion of the EG-PES

under the standards highlighted in the literature allowed us to justify that the survey complied with good standards, identified and attempted to minimize potential survey errors where possible. Given that the EG-PES followed best practices in sampling design, it is fair to assume that the final sample represents the underlying population and that the empirical analysis conducted under chapters 4 and 5 can be generalized to the population of grades 1, 3 and 6 students in Equatorial Guinea. Overall, because of the sampling approach adopted, chapters 4 and 5 are able to make inferences to the population that was the target of the investigation.

Although some potential small sources of survey error though interviewer effect were identified, such as differences in interviewer ethnic group and potential lack of standardization in applying the questionnaires, overall the EP-GES applied and followed good standards to minimize error in each phase of the survey. There could be a social desirability bias concern on the use of self-reported variables in the analysis under chapters 4 and 5, specifically the use of teachers and principals self-reported information on how frequently they meet to collaborate on pedagogical topics. However, EG-PES interviewers adopted practices to minimize potential bias: (i) they disclaimed, before the start of the interview, that the information provided by the respondent would be strictly confidential and protected by codes; and (ii) made clear to respondents that he or she would not receive any personal benefit as a result of the survey. In addition, we believe teachers and principals did not have any incentives or unobserved motives to over report their collaboration practices (which are the focus of chapters 4 and 5), since the country does not provide any incentive to teachers or principals who collaborate more frequently. Also, at the time of the survey, collaboration was not a practice incentivized or promoted by local or national policies. Nonetheless, we acknowledge the limitations of relying on self-reported data and the possibility of respondents over-reporting collaboration, which would ultimately impact the research findings. Finally, the survey followed ethics considerations to protect human subjects that took part in the research. Overall, it is reasonable to assume that the data yielded by EG-PES is accurate and reliable.

References

Alcser, K., Clemens, J. (2010). Cross- Cultural Survey Guidelines. Chapter X. Interviewer Recruitment, Selection, and Training. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan. Retrieved from: http://ccsg.isr.umich.edu/iwerselection.cfm

Berazneva, J. (2014). Audio Recording of Household Interviews to Ensure Data Quality. *Journal of International Development*, *26*(2), 290–296.

Bradburn, N. M., Sudman, S. (2011). In *Measurement Errors in Surveys*, P. Biemer, and others, eds. New York: John Wiley and Sons, pp. 29-40.

Blair, J., Czaja, R. F., & Blair, E. A. (2013). *Designing Surveys: A Guide to Decisions and Procedures*. SAGE Publications.

Billiet, J., & Loosveldt, G. (1988). Interviewer training and quality of responses. *Public Opinion Quarterly*, *52*(2), 190-211.

Campanelli, P. (2008). Testing survey questions. In: de Leeuw, E. D., Hox, J. J., & Dillman, D. A. (Eds), *International handbook of survey methodology* (pp 176 -200). New York, NY: Taylor & Francis Group/Lawrence Erlbaum Associates.

Cochran, W., Mosteller, F., & Tukey, J. (1954). Principles of Sampling. *Journal of the American Statistical Association,49*(265), 13-35.

de Leeuw, E. D., Hox, J. J., & Dillman, D. A. (2008). The cornerstones of survey research. In: de Leeuw, E. D., Hox, J. J., & Dillman, D. A. (Eds), *International handbook of survey methodology* (pp 1-18). New York, NY: Taylor & Francis Group/Lawrence Erlbaum Associates.

Fowler, F. J. (1995). *Improving Survey Questions: Design and Evaluation*. Sage Publications: Newbury Park, CA.

Fowler, F. J., Mangione, T.W. (1990). *Standardized survey interviewing; minimizing interviewer-related error*. Sage Publications: Newbury Park, CA.

Fowler, F.J. (1991). Reducing interviewer-related error through interviewer training, supervision and other means. In *Measurement Errors in Surveys*, P. Biemer and others, eds. New York: John Wiley and Sons, pp. 259-275

Groves, R. M. (2004). *Survey Errors and Survey Costs*. New York: John Wiley.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2011).

Survey Methodology (2 edition). Wiley.

Kasprzyk, D. (2005). Measurement error in household surveys: sources and measurement. In *Household Sample Surveys in Developing and Transition Countries*. New York: United Nations Publications, pp. 171- 196.

Kish, L. (1949). A procedure for objective respondent selection within the household. *Journal of the American statistical Association*, *44*(247), 380-387.

LaRoch, S., Joncas, M., Foy, P. (2016). Sample Design in TIMSS 2015. In Martin, M. O. (Ed,), *Methods and Procedures in TIMSS 2015* (3.1-3.38). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/publications/timss/2015-methods.html

Lepkowski, J. (2005) Non-observation error in household surveys in developing countries. In *Household Sample Surveys in Developing and Transition Countries*. New York: United Nations Publications, pp. 200- 230.

Lohr, S. L. (2008). Coverage and sampling. In: de Leeuw, E. D., Hox, J. J., & Dillman, D. A. (Eds), *International handbook of survey methodology* (pp 97-112). New York, NY: Taylor & Francis Group/Lawrence Erlbaum Associates.

RTI International (2009). *Early Grade Reading Assessment toolkit*. Prepared for the World Bank, Office of Human Development, under Contract No.7141961. RTI International, Research Triangle Park, NC. Retrieved from: https://www.eddataglobal.org/documents/index.cfm?fuseaction=pubDetail&id=149

Ustun, T. B., Chatterju, S., Mechbal, A., Murray, C. J. L. (2005). Quality assurance in surveys: standards, guidelines and procedures. In *Household Sample Surveys in Developing and Transition Countries*. New York: United Nations Publications, pp. 200- 230.

Appendix 2.A. Kish Tables for Random Selection of Students and Classrooms

Table 2.A.1: Kish Selection Table for Grades with More Than Two Classrooms

Obs	Eligible	Classroom	sch1	sch2	sch3	sch4	sch5	sch6	sch7	sch8	sch9	sch10	sch11	sch12	sch13	sch14	sch15
1	3	1	B	B	C	A	A	C	A	A	A	B	C	A	A	A	B
2	3	2	A	A	A	C	B	A	C	C	C	C	A	C	C	C	A
3	4	1	B	D	D	D	D	C	A	B	A	B	A	B	B	A	B
4	4	2	A	B	C	A	B	B	B	D	C	A	B	C	C	B	A
5	5	1	E	E	C	C	E	E	C	A	C	D	D	D	D	E	B
6	5	2	B	D	E	A	C	C	A	C	A	B	B	E	B	D	C
7	6	1	B	C	E	B	F	F	D	B	D	A	C	C	F	A	A
8	6	2	E	E	B	A	D	B	E	C	B	B	B	B	D	B	D
9	7	1	G	B	A	A	A	F	D	D	E	E	B	A	C	B	E
10	7	2	C	D	C	B	B	C	A	B	G	G	C	B	A	E	B
11	8	1	H	G	B	A	B	C	F	A	C	C	A	F	C	C	E
12	8	2	G	D	G	D	E	D	H	D	B	F	G	H	H	G	D

Instructions on using the Kish selection table

1. Identify how many groups are in each of the grades selected for the survey. Only use the Kish selection table if there are more than 2 groups per grade.

2. Identify the column referring to the school. For example, if it is the first school surveyed, it would be "esc1". If it is the second school surveyed, it would be "esc2", and so on.

3. Find the two rows for the number of eligible groups (classes) in the school for each selected grade. For example, if the school has groups A, B and C for first grade, it has 3 eligible groups.

4. Find where the column and two rows you've identified intersect.

5. The letters in the cells for rows 1 and 2, where the column and row meet, are the groups chosen to proceed with selecting students.

Table 2.A.2: Random Start Student Selection Kish Table

Obs	Interval	sch1	sch2	sch3	sch4	sch5	sch6	sch7	sch8	sch9	sch10	sch11	sch12	sch13	sch14	sch15
1	2	2	2	2	2	1	2	1	2	1	1	2	1	1	1	2
2	3	3	1	1	1	3	3	1	3	1	2	1	2	2	3	3
3	4	1	4	3	2	4	4	4	2	3	3	2	1	3	2	4
4	5	4	3	5	1	2	4	2	1	1	1	5	4	3	5	3
5	6	5	5	2	2	3	5	2	4	2	6	4	6	3	3	1
6	7	1	5	2	6	6	1	3	3	2	2	6	6	4	4	6
7	8	7	6	7	7	5	6	4	1	2	4	4	6	2	3	7
8	9	8	2	3	6	2	3	6	2	5	8	8	3	3	4	4
9	10	10	5	9	3	5	7	7	1	7	9	3	5	9	4	10

Instructions for Student Selection

1. Count the total number of students in the classroom chosen for the survey (for example, 24 students)
2. Determine the total number of students who will be interviewed (8 if there is only one group in the selected grade or 4 if there are two groups)
3. Divide the total (24) by the total number of students to be interviewed (for example, 8). In this example, 24÷8 = 3, is the sampling interval.
4. Select a random starting point between 1 and 3 with the help of the Kish table as follows:
    a. Identify the column referring to the school. For example, if it is the first school surveyed, it would be "esc1". If it is the second school surveyed, it would be "esc2", and so on. In this example, we have esc1.
    b. Find the row that corresponds with the sampling interval, in this case 3.
    c. Find where the column and row intersect.
    d. The number where the column and row intersect is the random starting point between 1 and 3. In this example, it would be 3.
5. According to the organization of students based on their seating in the class, select 3 students in this example: 3, 6, 9, 12, 15, 18, 21, 24.
6. If a student chooses not to participate, select the student that follows.

Appendix 2.B. Oral Consent Scripts

Consent Script for Principal

General Instructions

It is important to introduce the research study, clarify what will be studied in the school during the visit, and explain the possible risks involved.

Verbal Consent

Read only the text in the boxes below clearly to the principal:

Introduction to research study

My name is _____. I work with the Program for Educational Development of Equatorial Guinea (PRODEGE), and we are doing a research study to assess teachers' performance and students' math and reading competencies in primary education. I would like to ask your permission to perform the study activities within your school and your consent for you to participate in this study through an interview. Your permission and your participation in the study is totally voluntary. I want to ensure that you understand the purpose, responsibilities associated with the study, and potential risks before you decide whether or not to participate. Please do not hesitate to ask me questions or for further explanation.

Participation of the principal and school in the study

Your school was selected randomly using a statistical sampling process. We request your permission to conduct the study in your school and your help in this study but you do not have to participate if you would rather not. Your name will not be mentioned in any part of the reports based on this research. The name of your school will be registered to allow access to School Census data.

If you agree to participate, I would like to ask you some questions regarding your activities in the school, including your interactions with the staff of the school and that of the MEC. We are not evaluating your performance as a school principal, but rather we are trying to understand the profile of primary school directors, the organization of the school, and types of support received. If any questions make you feel uncomfortable, it is not necessary to respond.

With your help, I would select students from the first, third and sixth grades at random to evaluate their reading and math skills. We will also ask the selected students questions about their school activities and families. The students who are selected may choose to participate in the research or not. No student in the school will be identified in the investigation: all of the information will be maintained using codes and it will be used solely for the purpose of research.

We will also observe teaching practices in first, third and sixth grade classes and administer an interview questionnaire. The teachers' names will be collected to track their participation in PRODEGE trainings. However, all of the information will be coded and will only be used for the purpose of research.

Possible Risks and Benefits

The primary risk of this study is the violation of the confidentiality, which means that your name could be associated with the information that you provide us. In the case that the data was obtained by someone outside of this study team, it is possible that the information could be utilized in detriment to the participants. We will work to ensure confidentiality to the extent possible in order to prevent wrongful access to this information. Although this study will not benefit you personally, we hope that our results will contribute to the knowledge about improving the quality of primary education in Equatorial Guinea and at the same time supporting the work of principals, teachers and other members of the education system.

Confidentiality

We will assign a numeric code to all the information that you provide us and keep it separate from your name. All the collected information will be stored in a secure place that we use for the research study. We will not use your name in any publication, nor will we cite your name in a way in which you could be identified.

For further information

If you have a question about this activity, please call Fernanda Soares, the main point of contact from FHI 360, the organization implementing this research study for PRODEGE. We can provide her contact information if needed.

This study has been reviewed and approved but the Institutional Review Board of FHI 360. If you have a question about how the information is being used or about your rights as a participant, you can contact the Protection of Human Subjects Committee (PHSC). We can provide the Committee's contact information upon your request.

Final permission

Do you allow the team to conduct the study activities in your school? Once again, you do not have authorize the study activities in your school.

If verbal permission is not obtained, proceed with the activities within the school. If verbal permission is not obtained, thank the principal and end all work in this school.

Final consent

Are you willing to participate in the study through an interview? Once again, you do not have to participate if you would rather not. Do you have any questions about your role in the research? Do you agree to volunteer as a participant in this research? May we begin?

If verbal consent is obtained, proceed with the interview. If verbal consent is not obtained, thank the principal and proceed with the other school activities.

Observation:

If the principal requests the contact information for Fernanda Soares or the Protection of Human Subjects Committee (PHSC), please provide the information below:

Fernanda Soares Contact Information
Email: fsoares@fhi360.org
Telephone number: 222.16.95.57

Protection of Human Subjects Committee (PHSC).
PO Box, Research Triangle Park, NC 27709, USA.
Telephone number:+1-919-405-1445
Email at PHSC@fhi360.org.

Consent Script for Teacher

General Instructions

It is important to introduce the research study, clarify what will be studied in the school during the visit, and explain the possible risks involved.

Verbal Consent

Read only the text in the boxes below clearly to the teacher:

Introduction to research study

My name is _____. I work with the Program for Educational Development of Equatorial Guinea (PRODEGE) and we are doing a research study to assess teachers' performance and students' math and reading competencies in primary education. I would like to ask your consent to participate in this study. Participation in the study is totally voluntary. I want to ensure that you understand the purpose, responsibilities associated with the study, and potential risks before you decide whether or not to participate. Please do not hesitate to ask me questions or for further explanation.

Participation of the teacher in the study

Your school was selected randomly using a statistical sampling process. We request your help in this but you do not have to participate if you would rather not. It is not a work requirement. Your name will not be mentioned in any part of the reports based on this research. The name of your school will be registered to allow access to School Census data.

If you agree to participate, I would like to observe your classes for a few minutes and then ask you some questions about them and your activities in the school, including your interactions with the staff of the school and that of the MEC. We are not evaluating your performance as a teacher, but rather we are trying to understand the profile of primary school teachers, the use of time in the classroom, types of interactions between teachers and students, and types of support received. If any questions make you feel uncomfortable, it is not necessary to respond.

Possible Risks and Benefits

The primary risk of this study is the violation of the confidentiality, which means that your name could be associated with the information that you provide us. In the case that the data was obtained by someone outside of this study team, it is possible that the information could be utilized in detriment to the participants. We will work to ensure confidentiality to the extent possible in order to prevent wrongful access to this information. Although this study will not benefit you personally, we hope that our results will contribute to the knowledge about improving the quality of primary

education in Equatorial Guinea and at the same time supporting the work of principals, teachers and other members of the education system.

Confidentiality

On data collection forms we will record your name. However, we will assign a numeric code to all the information that you provide us and keep it separate from your name. All the collected information will be stored in a secure place that we use for the research study. We will not use your name in any publication, nor will we cite your name in a way in which you could be identified. Your name will only be used to connect you to those of your students who participate in the survey and to track information regarding PRODEGE trainings you have participate in.

For further information

If you have a question about this activity, please call Fernanda Soares, the main point of contact from FHI 360, the organization implementing this research study for PRODEGE. We can provide her contact information upon your request.

This study has been reviewed and approved but the Institutional Review Board of FHI 360. If you have a question about how the information is being used or about your rights as a participant, you can contact the Protection of Human Subjects Committee (PHSC). We can provide the Committee's contact information upon your request.

Final consent

Are you willing to participate? Once again, you do not have to participate if you would rather not. Once we begin the questionnaire or the observation, of you prefer not to answer a certain question or that we stop the observation that is fine. Do you have any questions about your role in the research? Do you agree to volunteer as a participant in this research? May we begin?

If verbal consent is obtained, proceed with the interview and observation. If verbal consent is not obtained, thank the teacher, and ask your supervisor for help selecting another teacher.

Observation:

If the teacher requests the contact information for Fernanda Soares or the Protection of Human Subjects Committee (PHSC), please provide the information below:

Fernanda Soares Contact Information
Email: fsoares@fhi360.org
Telephone number: 222.16.95.57

Protection of Human Subjects Committee (PHSC).
PO Box, Research Triangle Park, NC 27709, USA.
Telephone number:+1-919-405-1445
Email at PHSC@fhi360.org

Consent Script for Students

General Instructions

It is important to establish a playful and relaxed rapport with the children to be assessed, via some simple initial conversation among topics of interest to the child. The child should perceive the following assessment almost as a game to be enjoyed rather than a serious situation. It is important to read ONLY the statements in the box aloud slowly and clearly.

Verbal Consent

Read the text below clearly to the child:

- Let me tell you why I am here today. I work with the Program for Educational Development of Equatorial Guinea (PRODEGE) and we are doing a research study to understand how children learn. We are visiting schools all around EG and selecting children to help us. You were picked by chance.
- We would like your help in this. But you do not have to take part if you do not want to. Your teacher and principal have agreed for the school to participate in the study but they will not be mad at you if you do not want to participate.
- If you agree to participate, I am going to ask you to read letters, words and a short story out loud, and then do some math. How you do on the tasks will NOT affect your grades.
- I will also ask you other questions about your family, like what language your family uses at home and some of the things your family has.
- I will NOT write down your name so no one will know your answers.
- Once again, you do not have to participate if you do not wish to. Once we begin, if you would rather not answer a question, that's all right, just tell me and we will go on to the next question or we can stop altogether
- Do you have any questions? Do you agree to volunteer as a participant in this research? May we begin?

If verbal consent is obtained, proceed with the test. If verbal consent is not obtained, thank the child and move on to the next child, using this same script.

Chapter 3. Measuring Student Learning and Socioeconomic Outcomes

Over the past decade international and national agencies have begun to emphasize the need to improve education quality rather than quantity in developing countries (Wagner et al., 2012). The term education quality, however, has been applied to a variety of aspects of students' educational experiences, including safe learning environments, adequate infrastructure, availability of learning resources, curricula that are responsive to students' need, instructional practices, competent teachers and student learning. For the purpose of this dissertation, student achievement is employed as the main measure of education quality outcomes. This choice is not, however, uncontroversial and the use of tests to measure student outcomes has drawbacks. Student achievement tests have been criticized for not capturing the breadth of student outcomes that individuals and society value and that therefore are central to education quality (Ladd & Loeb, 2013). Tests often fail to capture important aspects of education quality highlighted by UNESCO (2005), such as education's role in encouraging learners' creative and emotional development, in supporting objectives of peace, citizenship and security, and in passing global and local cultural values. Tests also often ignore important skills for future outcomes, such as the ability to work effectively in groups, to empathize and to communicate effectively. This may be partially explained by the fact that many of these outcomes are not as easy to measure with currently available instruments (Ladd & Loeb, 2013). In summary, tests may be highly reliable measures of student achievement in the tested content, but it is important to keep in mind that they are unlikely to measure the full set of education goals (Ladd & Loeb, 2013).

Student achievement is the main dependent variable in two key research questions proposed by this dissertation and addressed in chapters 4 and 5 respectively:

1. Does teachers' level of collaboration affect student achievement?
2. Does principal's instructional support to teachers contribute to student achievement?

The choice of student achievement as measures of quality is based on recent research showing that children, especially in developing countries, are failing to achieve the expected competences for their grades. The 2009 EFA monitoring report (UNESCO, 2010) estimates that of the world's 650 million primary school age children, at least 250 million are not learning the basics in reading and mathematics. This situation is more acute in Sub-Saharan Africa, where only about 40% of the children reach grade 4 and achieve minimum learning standards in reading (UNESCO, 2010). Measuring student learning, however, it is not an easy task, and some controversy remains on how to design effective measurement instruments.

In what follows, this chapter provides an overview of measurement approaches to student learning and student socioeconomic status. First, we review student learning assessments in the developing

world and standards for quality of effective assessments. The specific case of a hybrid assessment used in the Equatorial Guinea Primary Education Survey (EG-PES) is explored in detail, with an analysis of its components and how it responds to the quality standards highlighted in the literature. Given that chapters four and five of this dissertation rely on data from the EG-PES hybrid assessment to construct the dependent variables student mathematic achievement and student language achievement, it is of key importance to assess the quality and soundness of this assessment. Given, the recognized role that family background and household socioeconomic status play in student learning (e.g., Sirin, 2005; Woßmann, 2003), we then compute a wealth index using data on household characteristics and durable asset ownership collected by the EG-PES. This index is used as a control variable in the empirical analysis undertaken in chapters four and five. Finally, we provide a descriptive overview for first and third grade students (which are the focus of chapters 4 and 5) performance in the communications and mathematics assessments.

3.1. Contextualizing: An Overview of Large-Scale Student Learning Assessments in Developing Countries

As the interest in educational assessments as a tool to measure student learning grew internationally, developing countries began to adopt and adapt international and other assessments for a variety of uses (Wagner et al., 2012). This section aims to review the objective, different types, categories and the applicability of learning assessments in developing countries.

Clarke (2012) defines assessment as "*the process of gathering and evaluating information on what students know, understand, and can do in order to make an informed decision about the next steps in the educational process*" (Clarke, 2012, pp. 6). The decisions based on the results of an assessment are related to the purpose of the assessment itself and may vary from designing of a system-wide program, to improving teaching and learning in the schools, to determining which applicants should be admitted to the next education level (Clarke, 2012). According to Braun & Kanjee (2006), assessments are essential to teaching and learning and are needed to monitor, evaluate, and improve the education system.

Clarke (2012), UNESCO (2012) and LMTF (2013) group student assessments into three major categories: continuous classroom assessments, examinations and large-scale assessments. Classroom assessments are carried out by teachers throughout the school year to support teaching and learning; they can be of formative or summative nature. Examinations are used with the purpose of making decisions about an individual student's progress through the education system or for credentialing learners. As such, they are high-stakes assessments that have a direct effect on the students tested. Large-scale assessments are designed for monitoring progress and providing policy-makers with information on system performance levels. These are usually low-stakes or no-stakes assessments with no consequences associated with student performance.

Large-scale assessments, which are the focus of this chapter, have been increasingly used by national and international agencies since the beginning of 1980s (Wagner, 2011). However, was not until the 1990s that developing countries significantly increased its participation and started to undertake their own large-scale assessments (Wagner, 2011). Two major international events contributed to the expansion of large-scale assessments in the developing world. In 1990, the Jomtien conference on "Education for all" demanded more accountability and systematic evaluations in developing countries. In 2000, the UNESCO Dakar Framework for Action called for achieving measurable learning outcomes, which should be monitored systematically (Wagner, 2011). Following these discussions and the widespread recognition that educational indicator systems must include not only inputs but also outputs (Braun & Kanjee, 2006), international agencies increased its funding and support to large-scale assessments in developing countries (Abadzi, 2012; Wagner, 2011).

With a focus on developing countries, Wagner, Babson & Murphy (2011) categorize large-scale assessments into four main types: national assessments, regional assessments, international assessments and hybrid. According to the authors, national assessments evaluate students in a national education system in order ascertain whether planned educational goals are achieved. Regional assessments measure student learning across a group of countries, usually defined by a geographic region or shared language[14]. Wagner, Babson & Murphy (2011) point out that regional testes have grown in popularity since the 1990s as part of an effort to expand large-scale assessments into developing countries. International assessments utilize standardized and uniform tests to assess learners in multiple countries with the aim to provide cross-national comparisons (Braun & Kanjeee, 2006; Lockheed, 2012). International assessments are usually conducted by international organizations and agencies[15]. Lockheed (2012) point out that the participation of low and middle-income countries in international large-scale assessments has increased from only a handful in the 1970s to nearly half of all developing countries by 2010.

A hybrid assessment is a new approach, which emerged in recent years, and involves the adaptation of tools used in the classrooms for formative assessments into surveys that collect information of learning, thus summative in nature (Clarke, 2012). Through the development of hybrid assessments, tools that before were only used inside of the classroom are adapted to document performance on the system level. Hybrid assessments focus on the specific needs of developing countries, which could potentially allow for greater frequency of implementation (Wagner, 2001; Wagner, Babson & Murphy, 2011). Since hybrid assessments do not have to follow the model for established regional

[14] Major regional assessments include: the Latin American Laboratory for Assessment of Quality in Education (LLECE), the Southern and Eastern African Consortium for the Monitoring of Education Quality (SACMEQ), and Program for the Analysis of Educational Systems of the CONFEMEN (francophone Africa) countries (PASEC).

[15] Major international assessments include: Trends in International Mathematics and Sciences Study (TIMSS), Progress in International Reading and Literacy Study (PIRLS), Program for International Student Achievement (PISA), and Monitoring Learning Achievement (MLS).

and international assessments, they often pay closer attention to local factors such as: population diversity, linguistic and orthographic, individual differences in learning and timeliness of analysis (Wagner, 2011). Hybrid assessments focus mostly on local contexts rather than on comparability across-contexts. Well-known examples of hybrid assessments include the Early Grade Reading Assessment (EGRA); the Literacy Boost Assessment, developed by Save the Children; the Annual Status of Education Report (ASER), implemented in India; and Uwezo, implemented in Kenya, Tanzania and Uganda.

As described in this section, in recent years there has been a significant growth in the number of countries, including developing countries, participating in learning assessments. These assessments, however, have varied purposes and designs: they can vary in the grades and age level tested, coverage of the population (sample or census), subjects or skills covered and frequency in which they are administered (Clarke, 2012). As Clarke (2012) points out, there is no ideal assessment system that works equally well in all contexts and the choice of an assessment will depend on various contextual factors and the goals of a given education system.

3.2. Early Grade Reading and Math Hybrid Assessments in Developing Countries

Early reading has been increasingly recognized as a key skill to succeed in school. As Gove and Wetterberg (2011) explain, unless children learn to read at an early age, they cannot absorb more advanced skills and content that rely on reading. Children who do not learn how to read in early grades may fall behind in later grades, as they cannot absorb printed information, comprehend written instructions or communicate well in writing (Gove & Wetterberg, 2011). Using longitudinal data from six different studies, Duncan et al. (2007) found that school-entry math and reading skills strongly predicted later school math and reading achievement. Similarly, in a study that followed first grade students until eleventh grade, Cunningham and Stanovich (1997) found that first grade reading ability was a strong predictor of 11th grade reading outcomes. Possible consequences of the inability to acquire early reading may include repetition and dropout, which strongly affect the efficiency of an education system. Research has also revealed that early literacy contributes to economic growth: a 10% increase in the share of students reaching basic literacy translates into a 0.3 percentage point higher annual growth rate for that country (Hanushek & Woessman, 2009).

Despite the potential impacts of early reading on school outcomes, most large-scale regional and international assessments undertaken in developing countries have focused on upper grades in primary education – grades four and above. They are typically administered as group-level testing in schools and assume students can read and write, requiring them to be skilled enough to complete a written examination independently (Wagner, 2011). When students score poorly, evaluators cannot determine if that is because they lack the knowledge tested by the assessment or because they lack basic reading and comprehension skills (Gove & Wetterberg, 2011). Therefore, this approach

is not appropriate to measure students at beginning reading, which is the case of students even after two or more years of schools in developing countries (Wagner, 2011).

In the context of these findings, international donors have called for the creation of measures to assess reading and math skills in early grades (Gove & Wetterberg, 2011). In response, hybrid assessments have been developed[16] to focus mainly on the period from first to third grade and properly assess reading and math competence much earlier than the other regional and international assessments. Focusing on early grades is possible as hybrid assessments are typically administered on a one-on-one basis and are largely oral assessments given to children (Wagner, 2011).

The development of hybrid assessments has been informed by research on early reading acquisition. Research has indicated that three main literacy skills contribute to reading achievement: phonological awareness, print knowledge and orthographic knowledge (Dubeck & Gove, 2015). The ability to identify sounds in words and to separate words into sounds is termed phonological awareness (Gove & Wetterberg, 2011). Print knowledge refers to alphabet letters recognition and knowledge of letter name and sound (Dubeck & Gove, 2015). As learners acquire a full understanding of the spelling system, they build an orthographic framework (Gove & Wetterberg, 2011), characterized by an understating of words in their written form. This includes the knowledge that certain sequences of letters compose words that represent spoken sounds (Dubeck & Gove, 2015). Initially students read letter by letter but with time they start to automatically identify entire words (Abadzi, 2012). Subsequently, learners progress to not only decoding letters into sounds, but also attributing meaning to written text (Gove & Wetterberg, 2011).

International hybrid assessments such as EGRA, are built under the assumption that learning to read in all alphabetic languages requires acquisition of similar foundation skills, although the importance of each of those skills may vary by language (Gove & Wetterberg, 2011). The assessments are broken down into subtasks that correspond to the different building blocks of reading acquisition. The foundational reading skills assessed by most hybrid assessments, including EGRA, ASER and Uwezo, include letter naming, reading simple words, sentences and a short paragraph, which is paired with comprehension questions (LMTF, 2013).

One measure used in some hybrid assessments is reading fluency, defined as the number of correct words read in one minute. Reading fluency is considered a measure of overall reading competence, as it captures the ability to translate letters into sounds, to unify sounds into words, to process connections, to relate text to meaning, and to make inferences to fill in missing information (Gove

[16] EGRA is the more well know example of hybrid assessment that has been created to meet donors demand and has been used in more the 65 countries since its creation in 2006 (Dubeck & Gove, 2015). The rapid expansion of EGRA is particularly explained by the support from the United States Agency for International Development (USAID), which account for roughly half of the applications to date (Dubeck & Gove, 2015).

& Wetterberg, 2011). Research has also shown that reading fluency is a prerequisite for comprehension: the faster one reads a message, the more time one has to put it into working memory, connect it with prior knowledge and thus understand it (Abadzi, 2012). Thus, working memory play a critical role in the relationship between reading fluency and comprehension. However, working memory may be more limited among lower income children in developing countries due to the impact of health problems (Abadzi, 2012).

The relationship between reading fluency and comprehension is particularly strong in early primary grades: beginning readers may not acquire much information from a text if they read slowly and struggle to retain the content in working memory (Abadzi, 2012). Simply stated, if a student reads too slowly, by the end of a sentence he/she may forget the beginning. The relationship between speed and comprehension weakens in more advanced grades, as vocabulary, background knowledge and literary analysis of narratives become more important than reading speed for comprehension (Abadzi, 2012; Fuchs et al., 2001).

Research in psychology and biophysics has recognized that for simple texts and limited prior knowledge, a minimum reading speed of seven words per twelve seconds is necessary in order to comprehend the text (Abadzi, 2012). This represents a minimum benchmark of about forty-five words per minute. Abadzi (2012) points out that across several languages and scripts 45-60 words per minute amount to eighty percent comprehension when vocabulary is known and point to automaticity.

Reading comprehension in hybrid assessments has been criticized as being memory tasks (Dubeck & Gove, 2015), although its instructions have differed. In some cases, after the reading passage is orally read, it is removed from the view of the student and the comprehensions questions are asked relative to what was read during one minute (Dubeck & Gove, 2015). In other cases, students are able to read the entire text past the one-minute mark and sometimes even to re-read a text while responding. While all approaches are acceptable, they measure different things. When the child is not allowed to re-read the passage, he/she is assessed on the ability to recall or to infer, depending on the type of question asked. When the written passage is returned to the child, he/she is also assessed on the ability to negotiate text and skim to locate the answer (Dubeck & Gove, 2015).

3.3. Standards of quality, validity and reliability

AERA, APA, and NCME (1999) define assessment quality as the psychometric quality, processes, and procedures used for the assessment activity. As Clarke (2012) points out, assessment quality is important because if an assessment is not sound, it may contribute to poor decision-making in regard to student learning and system quality.

The two most important characteristics of well-designed assessments are validity and reliability. Through an analysis of validity and reliability, it is possible to determine how well the overall assessment performs in evaluating what students know and can do. Waugh and Gronlund (2012) define validity as the appropriateness and meaningfulness of the inferences we can make from assessment results. Three major categories accumulate evidence to support the validity of an assessment results.

The first category is content-related evidence of validity, which refers to the adequacy of the sampling of tasks or items included in an assessment. Since an assessment is always a sample of the many tasks that could be included in it, content validation determines if the sample included is representative of the larger domain of tasks it is supposed to represent (Grenaney & Murray, 2012). Waugh and Gronlund (2012) highlight that content related evidence of validity can be achieved by following a systematic procedure for specifying and selecting the sample of test items, constructing high quality items and arranging the test for efficient administration and scoring (Grenaney & Murray, 2012).

Criterion-related evidence refers to the degree to which assessment results are related to some other valued measure called criterion (Grenaney & Murray, 2012). Two types of studies are used to obtain criterion-related evidence. The first type, called predictive studies, use test results collected first to predict future performance on criterion data collected at a later point in time. The second type refers to concurrent studies, which use test performance to estimate current performance on some criterion. Concurrent studies might be useful to check the results of a newly constructed test against some existing test that has validity evidence supporting it (Grenaney & Murray, 2012).

Construct-related evidence refers to whether an assessment adequately measures the underlying (latent) constructs in question (Kazdin, 2003). Construct-related evidence can be assessed through many different methods and is often an iterative process. Waugh and Gronlund (2012) point out that the construct related category encompasses a variety of types of evidence, which may include:

1. A description of the theoretical framework that specifies the nature of the construct to be measured;
2. A description of the development of the test and any aspects of measurement that may affect the meaning of the test score;
3. The pattern of relationship between the test scores and other significant variables (e.g., high correlation with similar tests).

The construct-related category of evidence is broad and involves several types of evidence that might be considered. Kellaghan, Grenaney, and Murray (2009), for example, describe three main types of evidence that assessments need to comply with to provide valid information on students' knowledge and skills. First, assessments should ensure that they provide an accurate representation

of the curriculum knowledge and skills or constructs they intent to measure. The authors point out that to secure adequate representation of a domain or construct, a test should contain an adequate number of items. With a small number of items it can be difficult to justify that a specific measurement objective can be achieved. Second, an assessment should assess knowledge and skills at a level that is appropriate for the students that will take it. This implies taking into account not only the standards highlighted in the curriculum, but also what is known of the actual students' achievements in schools. This is particularly true for developing countries, where often student achievement is not aligned with curriculum expectations. Third, students' performance should not be determined by their competences in domains other than the one that the test was designed to assess. As an example, Kellaghan, Grenaney and Murray (2009) point out that students' achievement in science or mathematics test should not contain so much language that performance on it depends on the ability to read rather than on the ability in science and math.

Reliability refers to the consistency of assessment results (Grenaney & Murray, 2012). An assessment is said to have high reliability if it is produces similar results under consistent conditions. This consistency of results indicates that they are relatively free from error (Grenaney & Murray, 2012). Reliability is necessary to obtain valid inferences from assessment results, as it provides the consistency of results that makes valid inferences possible (Grenaney & Murray, 2012). Reliability is typically reported by means of a reliability coefficient. Reliability coefficients can be obtained by a number of different methods, such as test-rest, equivalent form, test-retest with equivalent forms and internal consistency. Each one of these methods measure a different type of consistency: consistency over time, over different samples of items, or over different parts of the test (Grenaney & Murray, 2012). Examples of coefficients include the Cronbach's alpha and the Kuder-Richardson estimate. The Cronbach's alpha (coefficient alpha) is a descriptive estimate of the interrelationships among a set of items (Kazdin, 2003). Cronbach's alpha indicates how reliably a set of items measures a particular underlying (latent) construct. Cronbach's alpha can range from 0 to 1.00, with higher numbers indicating a more reliable measure. This coefficient is a function of the number of items used, such that assessments with fewer items generally have lower reliability estimates. A typical cut-off for acceptable reliability on a scale used for research purposes is .70 or higher (Salvia, Ysseldyke, & Bolt, 2010). Measures with reliability coefficients below this standard may be used but should be interpreted with caution.

3.4. Standards of Quality in the Student Learning Assessment in the Equatorial Guinean Primary Education Survey (EG-PES)

One of the main purposes of the EG-PES was to collect information on student learning on grades 1, 3 and 6 of primary education, which represent the beginning, middle and end of the primary education cycle. EG-PES team previous knowledge of the country education context suggested that

children in grades 1 through 3 of primary education had only beginning reading skills. Given the likely low readings levels in the country, as well as challenges to apply group-level assessments in developing countries (described in section XII of the present chapter), the EG-PES decided to develop and implement hybrid assessments. Because hybrid assessments are administered orally and on a one-on-one basis, the EG-PES team considered it would more reliably measure beginning reading and math skills, especially for students in grades 1 and 3. The present section provides a description of the learning assessments developed and applied as part of the EG-PES and reviews the conceptual framework in communications and mathematics that guided the assessment development decisions. Data from this assessment is used to compute the student achievement variables in mathematics and language employed in the empirical analysis under chapters 4 and 5.

Communication and Mathematics are key competencies that have been prioritized both globally and by the Ministry of Education in Equatorial Guinea (EG). In EG, the skills within each of these competency areas are identified and sequenced in a curriculum map. In the first cycle of primary school (grades 1-3), the teaching and learning process focuses almost exclusively on communication (oral language, reading and production of texts) and early numeracy. Since many children in EG arrive at school with limited skills in Spanish, oral language and beginning reading skills are emphasized in the curriculum mapping for first grade. The goal of the curriculum is for all children to be reading and understanding at least simple texts by the end of first grade. During the second cycle (grades 4-6), these competencies continue to be taught as subject areas and they are integrated with other subject areas such as science and social studies. For students in fourth to sixth grade, the curriculum focus shifts from "learning to read" to "reading to learn". In order for students to use reading as a tool for learning, they need reading fluency and reading comprehension, both globally recognized indicators of literacy for primary school students.

The assessment instruments used in the EG-PES focused on the prioritized competence areas (Communications and Mathematics). To develop the communication assessments, a team of local educators and international consultants drew on the international research on reading achievement reviewed in section 3.2, which shows that literacy skills in alphabetic languages develop along predictable patterns: phonological awareness, print knowledge and orthographic knowledge (Dubeck & Gove, 2015). The team also built on existing research on measuring reading to develop assessment items corresponding to the skills outlined in the EG curriculum map at each grade level. In this process, assessment strategies identified by Marie Clay (1979) and other hybrid assessments were revised and adapted.

The result was a research-based grade-level collection of individual subtasks that measure foundational literacy skills as well as reading and writing skills. The assessment included subtasks that are commonly used by Uwezo and ASER, such as letter naming, reading simple words, and a

short paragraph, which is paired with comprehension questions. However, additional measures were also included, such as subtasks that measure phonological awareness, and subtasks that measure reading competence by focusing on reading fluency and reading accuracy. Overall, measures included in the assessment aimed to provide insights regarding the pre-literacy skills of non-readers as well as the literacy skills and reading comprehension of readers. All materials used in the instruments (e.g., reading passages, dictation sentences) came directly from the EG curriculum and were adapted from local textbooks.

The term "communications" was used to name the assessment rather than literacy (which is a more commonly used term internationally) for two main reasons: 1) to follow the curriculum structure in Equatorial Guinea, which has "communications" as key competency in the curriculum and an official subject; 2) while this assessment assesses most literacy skills included in literacy assessments (such as reading, reading comprehension and writing), it also goes beyond to include skills such as listening comprehension and auditory discrimination. While the total score of the communications assessment may not be comparable with other countries, it is possible to compare student performance on specific indicators, such as Words Read Per Minute (also called Oral Reading Fluency in other assessments) and percent of non-readers. Sub-section 3.6 presents an overview of the outcomes on these key indicators as well as a brief comparison with other countries.

The research team also developed a collection of math assessment items corresponding to the competences outlined in the EG curriculum map for each grade level. Based on item analyses of the pilot results and consistency with the curriculum map, math items were selected for each grade level.

Pilot testing

As discussed in Chapter 2, data collection instruments may influence the respondent's answer through several channels: wording of the questions, questions length, the difficulty of the words used, open versus closed questions, the position of the question, among others (Bradburn & Sudman, 2011). With the objective of improving the assessment design and minimizing respondent-related errors, pretests using draft versions of the instrument were conducted. The EG-PES team piloted draft versions of the assessments in six schools in June, 2014, on a sample of 99 grade one, three and sixth students. The interviewer debriefing technique explained in Chapter two was applied: interviewers wrote a report detailing their observations, problems with items and making suggestions for modifications. As a result of the pilot, student instructions, examples, and test items that students had trouble understanding were refined and improved. The language of the assessment was adapted to include local expressions and vocabulary.

A psychometric analysis was also conducted with the pretest data to identify problematic items and subtests and to guide the research team in improving the reliability of the assessment. The reliability coefficient (Cronbach alpha) for the mathematics assessment for all three grades was above 0.8.

Specifically, it was 0.94 for the first grade math assessment with 41 items; 0.91 for the third grade assessment with 58 items; and 0.81 for the sixth grade assessment, which contained 31 items. The coefficient alpha for the communication assessments varied by grade and sub-task, ranging from 0.59 to 0.99. This lower alpha in certain subtasks, however, was likely due to the low number of items. In addition to the psychometric analysis, the time to complete the assessment as well as the level of difficulty was taken into consideration. For example, the mathematics assessment for third grade had a high number of zero scores and seemed very long and difficult to the children. The assessment was shortened, by cutting out some of the items, and other items were adapted to match the level of the students' learning.

Based on the pilot findings, the draft versions were revised, adjusted and reapplied in the fall of 2014 on a small sample of six students in order to further refine the items and to assess the testing times needed for each section and grade level. The assessment measures described in this report were all individually administered so estimating testing time and trimming the tests to fit within the available time in each school was required.

A variety of reading passages (at least three per grade level) were included in an initial pilot of potential achievement instruments in June of 2014. Reading comprehension questions were developed for each passage to assess both literal and inferential comprehension. Based on the results of the pilot, passages were selected for each grade level that were neither the hardest nor the easiest but rather seemed to be representative of grade level expectations.

A final review of the testing materials was conducted as part of the training of data collectors. Although the materials had been reviewed multiple times during the training process by local educators and as part of the pilot administrations , the data collectors identified terminology that varied by region and they identified preferred wording. Prior to the baseline administration, the manuals, protocols and scoring guidelines were further refined to reflect this input.

Description of the final instrument

The final assessment instruments included subtasks designed to be grade-level appropriate and to cover a range of skills. The sub-tasks were administered individually in an interaction between the interviewer and the student. To minimize interviewer-related errors, as explained in Chapter two, interviewers were trained on standardized assessment administration protocols.

The final communications assessment for first grade include subtasks that measure literacy and comprehension skills as well as some of the foundational skills needed for reading acquisition. Pre-literacy skills subtasks in the instrument include: letter identification, auditory discrimination, and listening comprehension. Another subtask included in the first-grade assessment is the number of words a child can read from a list of 30 "most frequently used words". This assessment strategy was

first identified by Marie Clay (1979) and subsequently adapted globally as a measure of children's familiarity with words that they are likely to encounter on a regular basis (Dowd, Friedlander, Guajardo, Mann & Pisani, 2013). For students in third and sixth grade, as explained previously the curriculum shifts the focus from "learning to read" to "reading to learn". In order for students to use reading as a tool for learning, they need reading fluency and reading comprehension, both globally recognized indicators of literacy for primary school students. Key sub-tasks for third and sixth grade assessment are oral reading fluency (words read correctly in one minute) and reading comprehension. Table 3.1 details the collection of subtasks used in the assessment and their corresponding grade level. In the math assessment, first grade items focus on basic numeracy skills such as counting objects or pictures and doing simple adding and subtraction calculations. A few items covering money and measurement were included. The instructions and word problems were read aloud to allow non-readers to demonstrate their math skills. In 3rd and 6th grade, the math items include more advanced calculations as well as measurement, geometry and beginning algebra. Many of the items require the child to solve problems embedded in scenarios from daily life in EG. At each grade level, for each skill area, multiple items were developed. Tables 3.2, 3.3 and 3.4 detail the math collection of subtasks by grade. When applying individual math and communication assessments, it is important to minimize as much as possible the data collection time. That is because the last child to take the assessment was exposed to a few more days of school than the first child, potentially affecting learning. In the case of EG, the assessments were applied during a 20-school days' period. The duration of the data collection was calculated based on technical and logistical needs. Even though the time span between the first and the last child to be assessed was small (20 days), it is important to acknowledge that the children interviewed in the last days of the data collection might have had a slight learning advantage over the first few children assessed.

Reliability

As mentioned in section 3.2 of the present chapter, the Cronbach's alpha indicates how reliably a set of items measures a particular underlying (latent) construct. Tables 3.2, 3.3 and 3.4 show that alpha coefficients for total mathematics in grades one, three and six was above 0.75, which indicates that the assessment is reliably measuring overall grade-level math skills. Some specific math subtask, however, had a low alpha, which is also probably due the few items included in the subtask (number of items detailed in Tables 3.2, 3.3 and 3.4). Disaggregated math results by subtask are interpreted with caution in this research, especially in the cases where alpha did not reach 0.70. As illustrated in Table 3.1, alpha coefficient in the different communications sub-tasks for each grade varied from 0.58 to 0.99. Although in some of the sub-tasks the alpha is less than the recommended 0.70 for research studies, the low alpha in these cases is likely due to the few items included in the subtask. The low alpha is mainly observed in the writing subtasks, and the listening comprehension subtask for first grade. The overall reliability of the overall communications' assessment is 0.78,

0.36 and 0.23 for first, third and sixth grades respectively. Reliability of the first-grade assessment is considered reasonable by literature standards, above 0.70. However, reliability for the overall tool is low for third and sixth grades, even though the subtasks perform well separately. This indicates that third and sixth grade communication assessments may not perform as one dimension.

To further analyze if the communications assessments perform well as an integrated tool, I look at the intercorrelations between the subtasks for each grade, as shown in tables 3.5, 3.6 and 3.7. The intercorrelations for the first-grade assessment suggest that the correlation between listening comprehension and all the other variables is low. The strongest correlation amongst all the subtasks is the correlation between oral reading fluency and reading comprehension (0.88). This finding is in lined with the literature reviewed in this chapter, which indicates that reading fluency is a prerequisite for comprehension (Abadzi, 2012). The correlations between oral reading fluency and reading comprehension are weaker in the third (0.51) and sixth (0.25) grade assessments. This finding is also in lined with the literature, which suggests that the relationship between speed and comprehension weakens in more advanced grades, as vocabulary, background knowledge and literary analysis of narratives become more important than reading speed for comprehension (Abadzi, 2012; Fuchs et al., 2001). The strong intercorrelations and high alpha coefficient for the overall communications assessment indicate that first grade instrument performs well as an integrated tool. The intercorrelation and overall alpha coefficients are lower for the third and sixth grade assessments, indicating that these two instruments do not perform so well as integrated tools.

Although Cronbach's alpha provides information about how consistent an assessment measures an underlying construct, this estimate does not assess whether a particular scale is unidimensional or multidimensional. The underlying structure of each communication assessment was examined using exploratory factor analysis (EFA). EFA is a data-driven method that demonstrates how items on an assessment are associated, and whether items form one or more latent components (factors) within the scale.

Table 3.1. Description and reliability of the communications assessment subtasks by grade

		First Grade		Third Grade		Sixth Grade	
COMMUNICATION SUBTASKS	**Description**	**No. of items**	**Reliability (Cronbach's alpha)**	**No. of items**	**Reliability (Cronbach's alpha)**	**No. of items**	**Reliability (Cronbach's alpha)**
Letter name/sound identification	Measures knowledge of letters name/sound correspondence. 10 upper case and 10 lower case letters are presented in random order.	20	0.96	NA		NA	
Auditory discrimination	Measures knowledge of the letter that represents the initial sound in a word. It is oral and has 10 sets of words and corresponding picture.	10	0.89	NA		NA	
Listening comprehension	Measures receptive language of one or more orally read passage with explicit and inferential questions.	10	0.59	15	0.71	NA	
Frequently used words	Measures the ability to identify individual commonly used words from grade level text. 30 words are presented.	30	0.99	NA		NA	
Oral reading fluency	Measures the ability to read grade-appropriate and below grade-level texts. It is timed to 60 seconds and scores reading rate and accuracy.	-	-	-	-	-	-
Reading comprehension Texts 1 and 2	Measures the ability to answer questions about the grade-appropriate text used for assessing oral reading fluency. Include explicit and inferential questions and look backs are allowed. Combines the reading comprehension scores from texts 1 and 2.	10	0.77	10	0.70	10	0.77
Writing	Dictation	10	0.59	20	0.70	20	0.58
Overall Assessment			0.78		0.36		0.23

Table 3.2. Description and reliability of the mathematics assessment subtasks for first grade

MATHEMATICS SUBTASKS	**Description**	**No. of items**	**Reliability (Cronbach's alpha)**
Numbers	Beginning math skills using concrete objects and written numbers	13	0.89
Addition	Computing (includes word problems)	09	0.80
Subtraction	Computing (includes word problems)	07	0.81
Money	Problem solving with money	04	0.63
Miscellaneous	Identifying and measuring shapes	06	0.69
Total Mathematics	***Total assessment items***	***39***	***0.93***

Table 3.3. Description and reliability of the mathematics assessment subtasks for third grade

MATHEMATICS SUBTASKS	**Description**	**No. of items**	**Reliability (Cronbach's alpha)**
Calculation/ operations	Computing (includes word problems)	10	0.65
Measurement/ geometry	Identifying and measuring shapes	06	0.38
Money and time	Problem solving with money or time	06	0.55
Algebra	Identifying missing numbers in calculations or equations	05	0.68
Total Mathematics	***Total assessment items***	***27***	***0.80***

Table 3.4. Description and reliability of the mathematics assessment subtasks for sixth grade grade

MATHEMATICS SUBTASKS	**Description**	**No. of items**	**Reliability (Cronbach's alpha)**
Natural Numbers	Computing and problems with natural numbers	7	.77
Whole numbers	Computing and problems with whole numbers	4	.42
Rational and Decimal Numbers	Computing and problems with rational and decimal numbers	9	.51
Geometry	Identifying and measuring shapes	7	.56
Statistics	Working with graphs and tables	3	.89
Total Mathematics	***Total assessment items***	***31***	***0.78***

Table 3.5. Intercorrelations Among First Grade Communication Subtasks (N=481)

Subtask	1	2	3	4	5	6	7
1 *Letter name*	--						
2 *Auditory discrimination*	.76***	--					
3 *Listening comprehension*	.25***	.34***	--				
4 *Frequently used words*	.64***	.70***	.25***	--			
5 *Oral Reading Fluency*	.50***	.56***	.20***	.84***	--		
6 *Reading comprehension*	.48***	.57***	.27***	.82***	.88***	--	
7 *Writing*	.46***	.54***	.20***	.70***	.66***	.65***	--

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

Table 3.6. Intercorrelations Among Third Grade Communication Subtasks (N=468)

Subtask	1	2	3	4
1 *Listening comprehension*	--			
2 *Oral Reading Fluency*	.19***	--		
3 *Reading comprehension*	.48***	.51***	--	
4 *Writing*	.28***	.50***	.49***	--

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

Table 3.7. Intercorrelations Among Sixth Grade Communication Subtasks (N=246)

Subtask	1	2	3
1 *Oral Reading Fluency*	--		
2 *Reading comprehension*	.33***	--	
3 *Writing*	.33***	.25***	--

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

Factor Analysis

General guidelines for EFA include examining eigenvalues, and factor loadings to determine the number of latent variables in an EFA solution and to examine how items hang together on the assessment. Eigenvalues are variances of the original components underlying the set of items in the factor analysis. The number of eigenvalues with a value greater than 1.00 is generally accepted as a guideline for the possible number of latent constructs underlying a scale (Hamilton, 1991; Raykov & Marcoulides, 2008). A scale with only one eigenvalue above this point suggests a unidimensional scale, whereas a scale with multiple eigenvalues greater than 1.00 indicates a potentially multidimensional scale. However, although EFA might reveal multiple eigenvalues, and in turn multiple underlying factors within a scale, the decision to retain eigenvalues also depends on the substantive meaning or interpretability of the underlying factors, which is another criterion for evaluating dimensionality (Hamilton, 1991).

Factor loadings represent the relative contribution of a particular item to an underlying factor, and are similar to correlation coefficients, ranging between -1.00 and 1.00 (Raykov & Marcoulides, 2008). Typically, a factor loading of ±0.30 or greater is used as a guideline for an acceptable factor loading, or

association between the item and the underlying factor (Raykov & Marcoulides, 2008). In the case of multiple factors, items may load above ±0.30 on one or more factors, called a cross-loading. In this case, a rotation can facilitate interpretability of factor loadings.

The first-grade communication assessment was compiled by seven subtasks. EFA reveals that the eigenvalue for the first factor is quite larger than the eigenvalue for the next factor (4.4 versus .96). Additionally, one single factor, the first one, accounts for 63% of the variation contained in the entire tool. Factor loadings (Table 3.8) show that all subtasks, except listening comprehension, load above 0.30, which suggests that the main factor that measures reading skills is a combination of all the individual skills. The results suggest that the assessment is unidimensional and seems to measure an intrinsic skill, which is early reading ability.

Table 3.8. Eigenvalue and Factor Loadings for First Grade Communication Assessment (N = 480)

Factor	Eigenvalue	Proportion of the total variance explained by each intrinsic factor	Factor loadings for each communication skill on Factor 1	
Factor1	4.39761	0.63	*Letter name*	0.36
Factor2	.962761	0.14	*Auditory discrimination*	0.39
Factor3	.745267	0.10	*Listening comprehension*	0.18
Factor4	.421626	0.06	*Frequently used words*	0.44
Factor5	.218574	0.03	*Oral Reading Fluency*	0.42
Factor6	.140974	0.02	*Reading comprehension*	0.42
Factor7	.11319	0.16	*Writing*	0.37

The third-grade communication assessment was compiled by 4 subtasks. Table 3.8 reveals one eigenvalue greater than 1.00, which is 4.39 (Factor 1) and one principal factor that accounts for 63% of the total variation. This proportion is just slightly below 70% threshold recommended by the literature and indicate that there is indeed one intrinsic important factor. Factor loadings show that all subtasks load above 0.30.

Table 3.9. Eigenvalue and Factor Loadings for Third Grade Communication Assessment (N = 467)

Factor	Eigenvalue	Proportion of the total variance explained by each intrinsic factor	Factor loadings for each communication skill on Factor 1	
Factor1	2.23658	0.56	*Listening comprehension*	0.42
Factor2	.861692	0.22	*Oral Reading Fluency*	0.49
Factor3	.500297	0.13	*Reading comprehension*	0.56
Factor4	.401428	0.10	*Writing*	0.51

The sixth-grade communication assessment was compiled by 3 subtasks only. Table 3.9 reveals one eigenvalue greater than 1.00, which is 2.24 (Factor 1) and one principal factor that accounts for 56% of the total variation. Although this proportion is not above 70%, as indicated by the literature, the proportion is

not extremely low either and may indicate that there is one intrinsic factor. Factor loadings show that all subtasks load above 0.30.

Table 3.10. Eigenvalue and Factor Loadings for Sixth Grade Communication Assessment (N = 247)

Factor	Eigenvalue	Proportion of the total variance explained by each intrinsic factor	Factor loadings for each communication skill on Factor 1	
Factor1	1.6024	0.53	*Oral Reading Fluency*	0.42
Factor2	.753358	0.25	*Reading comprehension*	0.49
Factor3	.644245	0.21	*Writing*	0.56

Validity

Assessment validity was a key consideration during the process of assessment development and administration. To ensure content validation of the mathematics assessment, a systematic procedure was adopted to select a representative sample of assessment items, as recommended by Waugh and Gronlund (2012) and Grenaney & Murray (2012). First, a collection of items was developed with the assistance of a local math expert to assess each of the main math competences highlighted in the curriculum map. Subsequently, the collection of items was pretested during the pilot. The pilot reports and results showed that the students from third and sixth grades were not familiar with some of the math topics included in the test or were too difficult for their level. As a result of the pilot, the research team revised and refined the items, so they were at a level that is appropriate for the students that would take the assessment. At minimum four items were selected for each math competence being measured. This procedure ensured that the sample of items included is representative of the larger domain of tasks it is supposed to represent and appropriate for the level of knowledge and skills of students. To ensure construct validity, the math assessment was administered orally so that students' performance is determined by the ability in math rather than on competences in reading.

The communications assessment was developed following international research and standards to assess reading and pre-literacy skills. To build content validity into the instrument, the research team prepared country-appropriate items for each section of the assessment and adapted materials (e.g., reading passages, dictation sentences) from local textbooks. In addition, curriculum experts examined the first draft of the instrument and made recommendations about the appropriateness of each item type for measuring the early reading skills of students. Following this review, the research team adapted instrument as necessary.

Criterion-related evidence was not assessed as part of this research, as other external measures of student learning are not available in Equatorial Guinea. As RTI International (2009) point out, this type of validity evidence is hard to collect in countries with few standardized measures of student learning outcomes.

However, extensive research in other countries has demonstrated that EGRA-type instruments show strong relationships (0.7 and above) to other tests that measure similar constructs (RTI International, 2009).

3.5. Socioeconomic Measures

The literature recognizes that family background and household socioeconomic status play an important role in student learning and achievement (Sirin, 2005; Woßmann, 2003). Specifically, socioeconomically disadvantaged students tend to perform worse on standardized measures of academic achievement compared to their wealthier peers (Noel & de Broucker, 2001; Perry & McConney, 2010; PISA, 2004; Sirin, 2005).

Given the importance to control for students' socioeconomic background in the analysis of the influence of teachers-principals collaborative practices on student achievement (chapters 4 and 5 of this dissertation), I compute a wealth index using household characteristics and durable asset ownership. The student interview questionnaire of the EG-PES included a set of questions on housing characteristics, household infrastructure and yes-no questions on household ownership of nine types of durable assets. There is evidence in the literature that data on household infrastructure, building materials, and ownership of certain durable assets can be used to reliably measure inequality in living standards in the absence of information on household income or consumption (McKenzie, 2003; Filmer & Pritchett, 2011). To compute an index that enjoys internal validity, I use a Principal Component Analysis (PCA) to generate weights of each of the housing characteristics or durable assets included in the index. Using household assets and consumption data for Indonesia, Pakistan, and Nepal, Filmer and Pritchett (2001) find that PCA provides plausible weights for an index of assets. The authors further conclude that a wealth index computed using PCA provides a reasonable estimate of wealth levels.

The wealth index estimated in this research is based on 14 items related to housing characteristics and household ownership of durable assets. The housing variables were dichotomized to indicate the existence or not of a certain characteristic. The household ownership of durable assets variables were already binary variables, as they collected yes or no questions regarding ownership (1=Yes, 0=No). Weights were defined by factor scores for the first component of the PCA. Subsequently, a wealth index variable was created based on the 14 asset indicators and their respective weights.

Table 3.11 provides the scoring factors of the first principal component of the index, which was made from two categories, housing characteristics and durable assets ownership. The first component accounts for 21 percent of the total variation across the 14 indicators. The scoring factors are positive for all indicators, suggesting that ownership of the durable assets (included in the index) and improved housing characteristic included in the index have a positive impact on the wealth indicator. These coefficients suggest the first

principal component is indeed providing a measure of wealth. Having electricity, a bathroom inside the house and owning a fridge, ventilator, and computer raise the overall index by the largest amount, while owning a motorcycle, a bicycle or a radio have the least impact on the index. Table 3.11 also provides the mean and standard deviation for each of the 14 indicators included in the wealth index. There is a wide range in levels of ownership. Very few households have durable assets, such as a motorcycle and a bicycle. Most household however have a television and a cell phone. While most households have access to electricity, only a few have piped water. In order to assess the internal validity of the wealth index proposed here, terciles of wealth were computed. Table 3.12 presents means for each of the asset indicators considered in Table 3.12, for students classified into each wealth tercile. Table 3.12 indicates that the first Principal Component Analysis methodology discriminates well between the rich and poor, in the sense that mean asset ownership differs markedly across each tercile. Students in the upper tercile show much higher levels of wealth than the poorest tercile. While most students' household in the upper tercile own a fridge (99.7%), only 43% of students' household in the poorest percentile do; 11.3 percent of poor students' households have brick wall, compared to 77.3 percent of the richest tercile.

Table 3.11. Principal Components and Summary Statistics for Asset Indicators

VARIABLES	Scoring Factor for 1st principal component			Summary Statistics	
	Housing only	Durables only	All indicators	Mean	Std. Dev.
Housing Characteristics					
House has electric lighting	0.3718		0.3327	0.875	0.330
House has bathroom inside	0.5539		0.3412	0.409	0.492
House has kitchen inside	0.4299		0.2615	0.747	0.433
House has piped water	0.3263		0.1734	0.051	0.220
House has brick walls	0.5134		0.3150	0.389	0.487
Durable Assets Ownership					
Owns a television		0.4370	0.3003	0.919	0.273
Owns a radio		0.0297	0.0058	0.620	0.485
Owns a fridge		0.5134	0.3889	0.790	0.407
Owns a ventilator		0.4798	0.3622	0.686	0.463
Owns a computer		0.3631	0.3106	0.326	0.469
Owns a cellphone		0.2695	0.2088	0.920	0.272
Owns a bicycle		0.0684	0.0681	0.189	0.391
Owns a motorcycle		0.0376	0.0185	0.098	0.296
Owns a car		0.3221	0.2399	0.490	0.500
All indicators asset index				2.0571	0 .7179
Eigenvalue for first component	1.8365	2.10605	2.9104		
Share of variance first component	0.367	0.233	0.208		
Number of variables used	5	9	14		

Table 3.12. Means by tercile of all indicator asset index

	Summary Statistics		
VARIABLES	Lowest	Middle	Upper
Housing Characteristics			
House has electric lighting	0.678	0.950	0.997
House has bathroom inside	0.100	0.256	0.878
House has kitchen inside	0.519	0.776	0.948
House has piped water	0.00752	0.170	0.119
House has brick walls	0.113	0.452	0.773
Durable Assets Ownership			
Owns a television	0.792	0.967	0.997
Owns a radio	0.593	0.614	0.653
Owns a fridge	0.432	0.941	0.997
Owns a ventilator	0.285	0.824	0.951
Owns a computer	0.0659	0.263	0.653
Owns a cellphone	0.812	0.955	0.992
Owns a bicycle	0.160	0.139	0.269
Owns a motorcycle	0.0969	0.0670	0.129
Owns a car	0.269	0.484	0.719
All indicators asset index	1.239273	2.112748	2.82669

3.6. Overview of First and Third Graders' Performance

In this section, we review the performance of first and third grade students, which are the focus of chapters 4 and 5 of this dissertation, in the communications and mathematics assessments.

Communications Assessment – Results for First Grade

Table 3.13 summarizes first grade results on key indicators from the communication assessment. A key indicator for first grade is whether the child can read at least one word. Two tasks were used to measure this indicator yielding slightly different results. In one task, the child is presented with a short story (60 words). The story uses vocabulary that is typical of first grade reading exercises in Equatorial Guinea and is comparable to the measures used internationally (Gove & Wetterberg, 2011). For this exercise, the child is asked to read the story aloud. If the child reports that he/she is unable to read or the child tries to read but is unable to read any of the first 5 words, the task is stopped and the child is identified as a non-reader. Based on this measure, 44.3% of the children could read at least one word and 55.7% could not read. Using weighting, this translates into approximately 8,343 (42.5%) first graders nationally who could read and 11,274 (57.5%) first graders who could not read at the end of first grade. Results from Equatorial Guinea are in line with EGRA research findings in Africa, which have shown that a high percent of students at the end of second grade cannot read a single word from a reading passage. Although the percent of non-readers

is usually high, there are marked differences by country. In Zambia, for example, 91% of the students could not read at the end of second grade, where in Liberia this number falls to 30%[17].

In another task, the child is presented with 30 words that are used frequently in early reading materials (for example, "las" and "uno"). On this task, 57.8% of the children could read at least one word and 43% could not. The mean number of words read correctly was 12.42. Using weighting, this translates into 11,158 (56.9%) of first graders nationally who could read at least one of the 30 words and 8,459 (43.1%) of first graders who could not read any of the 30 words at the end of first grade. Figure 3.1 indicates how the scores were distributed based on the weighted sample—the largest group is non-readers but the next largest group (28 %) refers to the children who read between 26-30 words.

Figure 3.1. Distribution of children's total scores on reading 30 frequently used words

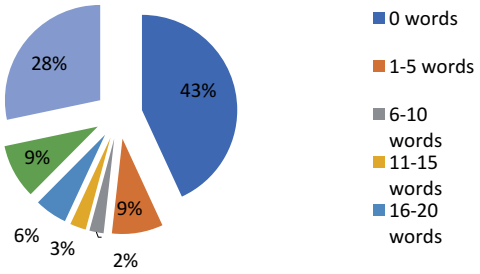


Table 3.13. Summary of First Grade Results on Key Indicators (Unweighted)

Indicator	Baseline Result	Number of observations
% that can read at least one word in frequently used words	57.8%	481
Mean score - Frequently used words read correctly	12.42	481
% that can read at least one word in connected text (Passage 1)	44.3%	481
Mean WRC Passage 1: Oral Reading Fluency (readers only)	32.82	213
Mean score Passage 1: Reading Comprehension (readers only)	2.91	213
Mean WRC Passage 2: Oral Reading Fluency (readers only)	27.71	197
Mean score Passage 2: Reading Comprehension (readers only)	2.69	197
Mean Score on Mathematics (Total)	20.04	474

[17] Assessing Early Grade Reading Skills in Africa, Retrieved from: file:///C:/Users/fsoares/Downloads/EGRA%20in%20Africa%20-%20Brief%20-%202011-08-19_FINAL_EndDateUpdated-Jun2014.pdf

The assessment was also designed to calculate Words Read Correctly (WRC) in one minute. When we analyze only the "readers" (i.e., only those students who read correctly at least one of the first 5 words in the story; n=213), there was wide variability in reading rate. The first story was short (60 words), included an illustration and used vocabulary that is emphasized in first grade (e.g., "Me llamo..."). If a student finished reading the text before 1 minute had elapsed, the time was noted and the WRC in one minute was extrapolated. On this first text, the average WRC was 32.82 (SD=18.74) with words per minute ranging from 2-109. On the second story, which was a little longer (100 words) with a wider range of grade level appropriate vocabulary, 16 fewer students could read the story and the mean WRC of those who could was 27.71 (SD=15.36). Of the 197 children who read both passages, the average WRC was 31.13 (SD=16.21). This rate translated to about one word every two seconds. Reading accuracy for readers was extremely high. On the first story, two thirds (65.8%) of the readers read all of the words correctly or made only one error. Similarly, on the second, slightly more difficult passage, 61.2% read all of the words correctly or made only one error. The data collectors were asked to describe why students didn't finish the stories. Of the students who read both passages, 75.5% were described as reading slowly, decoding each word, whereas 24% were described as reading without much difficulty but carefully so as not to commit errors.

Students who read the passages (i.e., they were able to read at least one of the first five words and they continued reading) were asked five questions about what they had read. They were encouraged to look back into the text if they were not sure of an answer. The average total correct for the first text was 2.91 (SD=1.68) and the average total for the second text was 2.09 (SD=1.26). The mean total for the 10 comprehension questions for those who read both passages was 5.17 (SD=2.54). These findings suggest that the readers were comprehending some but not all of what they read. Since look backs were allowed, this result also reflects the ability of the child to negotiate text and skim to locate the answer. The results show a high correlation between oral reading fluency and reading comprehension (0.88), which is in line with the literature indicating that reading fluency is a prerequisite for comprehension (Abadzi, 2012).

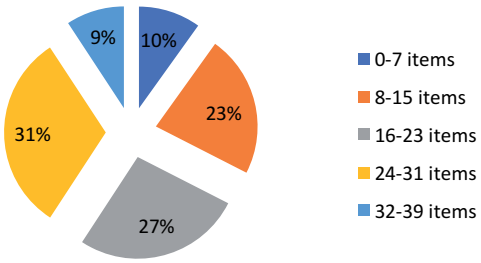
There were three writing tasks for first graders. Initially, all of the children in the sample were asked to write their names; 73.6% of the children were able to write at least their first name and 26.4% of the children could not. The children who could not write their names tended to be the non-readers (85.7% were non-readers). Children were then asked to write as many words as they could in 3 minutes. When the time had elapsed, they were asked to read the words they had written. Only words that were written and read correctly were counted as correct. The total number of words written correctly ranged from 0-30, the mean was 4.42 (SD=4.35) and 91.1% of the children wrote between 0-10 words. The third writing task was dictation. This task was only applied with the 214 children who had been able to write words. Two simple sentences were dictated and students' writing was scored on 5 criteria; 1 point was award for partially correct responses

and 2 points were awarded for completely correct. Students exhibited the most difficulty with punctuation (e.g., neglecting to end the sentence with a period). The average score was 4.94 (SD=2.97), out of 10.

Mathematics Assessment – Results for First Grade

The total math score is one of the key indicators and it is based on the sum of correct items out of 39 items. Items are distributed across a variety of topics in mathematics. Valid scores were available for 474 children in the sample and the mean score was 20.04 (SD=8.80). Based on the weighted sample, the mean was 20.06 (SD=8.83). Figure 3.2 illustrates the distribution of scores based on the weighted sample.

Figure 3.2. Distribution of first grade students' total math scores



Communications Assessment – Results for Third Grade

Table 3.14 summarizes results on key indicators for the third-grade sample. 439 (93.6%) of the 469 students evaluated could read at least one of the first five words in a brief text. Using weighting, this translates into 13,914 (93.2%) third graders nationally who could read at least one word and 1,011 (6.8 %) third graders who could not read one of the first 5 words in a text at the end of third grade. There was wide variability in reading rate. WRC ranged from 0-188 words per minute on the first text and 0-203 words per minute on the second text**.** The mean words per minute (including all students and both texts) was 53.57 words per minute (weighted mean for all students, both texts was 52.86). The median WRC was 51 words per minute or just under one word per second.

By contrast, there was little variability in reading accuracy. The majority of readers made no errors (37% on the first text and 42% on the second text made no errors) or very few errors (on both texts, more than 90% made four or fewer errors when reading aloud the selected texts).

Reading comprehension was significantly related to reading rate. Students who read more quickly also performed better on the questions about what they had read. Overall reading rate accounted for 24% of the

variance in reading comprehension. Worth noting is that students tended to perform well on the literal questions and they evidenced the most difficulty with the inferential items in the second (social studies) text. The high reading accuracy and low performance on inferential comprehension questions might indicate that there in an emphasis placed on reading correctly in classroom without focusing on the meaning of the text.

Table 3.14. Summary of Third Grade Results on Key Indicators (Unweighted)

Indicator	Baseline Result	Number of observations
% that can read at least one word in connected text	93.6%	469
Mean WRC-Oral Reading Fluency (both texts, all students)	52.69	469
Mean Score Reading Comprehension (both texts, all students)	4.99	469
Mean WRC-Oral Reading Fluency (both texts, readers only)	56.29	439
Mean score Reading Comprehension (both texts, readers only)	5.40	439
Mean Score on Mathematics (Total)	11.5	463

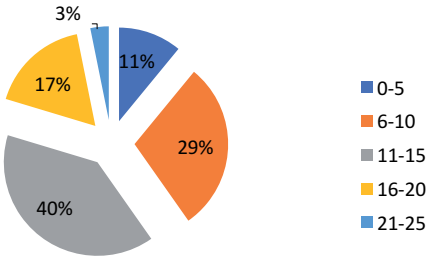
There were three additional communication tasks that provide insights into the skills of third grade students. Oral language was assessed by first asking students to comply with multistep instructions presented orally (part 1) and then asking the students to listen to two short stories presented orally and, after each story, to respond using complete sentences to questions about what they had heard (parts 2 and 3). Most students (62.9%) had no difficulty responding correctly to all 5 items in the first part however a small percentage of the children (6.8%) were unable to respond correctly to even half of the items. In the next two parts, there was wide variability in student performance however the percentage of children with very limited oral language comprehension persisted. For example, with a possible total equal to 20 points, 9% of the children obtained scores of 0-3 and almost 1/5th (19%) had scores between 0-5. These findings provide evidence that students' oral language fluency (receptive and expressive) in Spanish cannot be assumed. In fact, overall performance in these oral language tasks was a significant predictor of reading comprehension (accounting for 24.3% of the variance) and, when combined with WRC, accounting for 40% of the variance in reading comprehension.

Two other tasks assessed students' production of text. On the dictation task, students tended to write clearly and to spell commonly used words correctly. They lost points for failing to capitalize the first letter of proper names and for not including punctuation (period, question mark). On the writing task, students tended to write complete sentences and to describe the four pictures that were presented. They lost points for failing to include a title (as per the instructions) and forgetting to include periods at the end of sentences.

Mathematics Assessment – Results for Third Grade

The total math score is one of the key indicators and it is based on the sum of correct items out of 3 items. Items are distributed across a variety of topics in mathematics. Valid scores were available for 463 children in the sample and the mean score was 11.50 (SD=4.51). Based on the weighted sample, the mean was 11.55 (SD=4.62). Figure 3.3 illustrates the distribution of scores based on the weighted sample. Third graders tended to perform relatively better on the Numbers and Operations items and the Algebra (identifying missing numbers) items however there were insufficient items in each subgroup to warrant further interpretation.

Figure 3.3. Distribution of third grade students' total math scores



3.7. Concluding remarks

Although some controversy remains on the literature on how to design effective measurement instruments for student learning, there are certain agreed standards for quality, validity and reliability that instruments should follow. This chapter provided an overview of measurement approaches to student learning and socioeconomic status. The measurement approaches adopted by the Equatorial Guinea Primary Education Survey (EG-PES) were explored in detail. The choice for hybrid learning assessments was justified as suitable for early grades and contexts that present low readings levels and challenges to apply group-level tests, such as in Equatorial Guinea.

The data collected through the EG-PES student hybrid learning assessment informed the construction of the variables student mathematic achievement and student language achievement, which are key dependent variables used in the empirical analysis under chapters 4 and 5 of this dissertation. Checking the reliability and validity of the learning assessment, which is used in two empirical chapters of this thesis, is particularly important to justify the accuracy of the empirical results. A discussion on the development and psychometric properties of the EG-PES instruments allowed us to justify the reliability and validity of the instruments used.

References

Abdazi, H. (2012). *Developing Cross-Language Metrics for Reading Fluency Measurement: Some Issues and Options*. Washington: Global Partnership for Education. Retrieved from http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2013/07/26/000356161_20130726155230/Rendered/PDF/797740WP0wpm0e0Box0379789B00PUBLIC0.pdf

American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME). (1999). *Standards for Educational and Psychological Testing*. Washington, DC: AERA.

Braun, H. I., & Project on Universal Basic and Secondary Education. (2006). *Improving education through assessment, innovation, and evaluation*. Cambridge, MA: American Academy of Arts and Sciences.

Bruns, B. & Luque, J. (2014). *Great teachers: how to raise student learning in Latin America and the Caribbean*. Washington, D.C.: The World Bank Group.

Clay, M. M. (1979). *Reading: The patterning of complex behavior.* Portsmouth, NH: Heinemann.

Clarke, M. (2012). What matters most for student assessment systems: A framework paper. Retrieved from https://www.openknowledge.worldbank.org/handle/10986/17471

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology*, *33*(6), 934–945.

Dowd, A. J., Friedlander, E., Guajardo, J., Mann, N., & Pisani, L. (2013). *Literacy boost cross country analysis results. Save the Children.* Retrieved from: http://www.meducationalliance.org/sites/default/files/literacy_boost_cross_country_analysis_results.pdf]

Dubeck, M. M., & Gove, A. (2015). The early grade reading assessment (EGRA): Its theoretical foundation, purpose, and limitations. *International Journal of Educational Development*, *40*, 315–322.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., … Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428–1446.

Filmer, D., and LH Pritchett. (2001). Estimating Wealth Effect Without Expenditure Data or Tears: An Application to Educational Enrollments in States of India. *Demography 38*, 115-132.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral Reading Fluency as an Indicator of

Reading Competence: A Theoretical, Empirical, and Historical Analysis. *Scientific Studies of Reading*, *5*(3), 239–256.

Gove, A., & Wetterberg, A. (2011). *The Early Grade Reading Assessment: Applications and Interventions to Improve Basic Literacy*. Research Triangle Park, NC: RTI Press.

Hamilton, L. C. (1991). *Regression with graphics: A second course in applied statistics.* Pacific Grove, CA: Brooks/Cole Publishing Company.

Hanushek, E., & L. Woessmann (2009). *Schooling, Cognitive Skills, and the Latin American Growth Puzzle*. Working Paper 15066. Cambridge, MA: National Bureau of Economic Research.

Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). Identifying effective classroom practices using student achievement data. *Journal of human Resources*, *46*(3), 587-613.

Kazdin, A. E. (2003). *Research design in clinical psychology, 4th edition.* Boston, MA: Allyn and Bacon.

Kellaghan, T., Grenaney, V., & Murray, S. (2009). *National Assessments of Educational Achievement Volume 5: Using the Results of a National Assessment of Educational Achievement*. Washington, D.C.: The World Bank.

Ladd, H. F., & Loeb, S. (2013). The challenges of measuring school quality: Implications for educational equity. In D. Allen & R. Reich (Eds.), *Education, democracy and justice*. Chicago, IL: University of Chicago Press.

LMTF (2013). Toward Universal Learning: *A Global Framework for Measuring Learning. Report No. 2 of the Learning Metrics Task Force.* Montreal and Washington: UNESCO Institute for Statistics and Center for Universal Education at the Brookings Institution.

Lockheed, M. (2012). Policies, performance and panaceas: the role of international large-scale assessments in developing countries. In: The debate on learning assessments in developing countries. *Compare: A Journal of Comparative and International Education*, 42(3), 512-518.

McKenzie, D. J. (2005). Measuring inequality with asset indicators. *Journal of Population Economics*, *18*(2), 229–260.

Raykov, T., & Marcoulides, G. A. (2008). *An introduction to applied multivariate analysis.* New York, NY: Routledge.

RTI International (2009). *Early Grade Reading Assessment toolkit*. Prepared for the World Bank, Office of Human Development, under Contract No.7141961. RTI International, Research Triangle Park, NC.

Retrieved from: https://www.eddataglobal.org/documents/index.cfm?fuseaction=pubDetail&id=149

Salvia, J., Ysseldyke, J. E., and Bolt, S. (2010). Assessment in special and inclusive education. (11th edition). Belmont, CA: Wadsworth.

Sirin, S. R. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research*, *75*(3), 417–453.

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2005). Education for All, the Quality Imperative (EFA Global Monitoring Report 2005). Paris, France: Author.

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2010). *Reaching the marginalized* (EFA Global Monitoring Report 2010). Paris, France: Author.

Wagner, D. A. (2011). Smaller, Quicker, Cheaper. *Improving Learning Assessments for Developing Countries. Education for All Fast Track Initiative Secretariat*. Retrieved from http://www.safedafed.org/seminar/document/Assessment/Wagner_IIEP%20Assessments.pdf

Wagner, D. A., Lockheed, M., Mullis, I., Martin, M. O., Kanjee, A., Gove, A., & Dowd, A. J. (2012). The debate on learning assessments in developing countries. *Compare: A Journal of Comparative and International Education*, *42*(3), 509-545.

Wagner, D. A., Babson, A., & Murphy, K. M. (2011). How Much Is Learning Measurement Worth? Assessment Costs in Low-Income Countries. *Current Issues in Comparative Education*, *14*(1), 3–21.

Waugh, C. K., & Gronlund, N. E. (2012). *Assessment of Student Achievement* (10 edition). Pearson.

Wößmann, L. (2003). Schooling resources, educational institutions and student performance: the international evidence. *Oxford Bulletin of Economics and Statistics, 65*(2), 117-170.

Chapter 4. Teachers' Collaboration and Student Achievement: An Instrumental Variables Approach[18]

4.1. Introduction

Geared towards achieving Education for All in 2015, many Latin American and Sub-Saharan African countries have made a tremendous progress in the past decades in expanding enrollment and in increasing years of schooling. However, the expansion of the quantity and availability of schools came at the cost of the quality of education (Hanushek, 1995). To meet the increased enrollment needs, there has been a massive recruitment of less educated and/or untrained teachers, which has led to concerns about the quality of education being provided (Orr et al., 2013). Improving teaching quality to ensure that all children in school actually obtain the skills and knowledge they are meant to acquire have become key objectives (UNESCO, 2010).

Together with the growing recognition of the need to improve teaching, there is also an increased awareness that the isolation teachers often face is not conducive to collaborative learning and improved teaching practices (DuFour, 2004; Hord, 1997; Plessis & Muzaffar, 2010). In developing countries, many teachers, particularly those in small schools, feel isolated and disconnected from their peers. If teachers in developing countries have the opportunity to participate in professional development, it is often a one-off training in a central location, not connected to the context or the on-the-ground reality they experience in their school communities.

In order to improve teaching quality and to promote a social and contextual approach to teacher development (Plessis & Muzaffar, 2010) some countries have recently resorted to alternative models of teacher professional development, such as Professional Learning Communities (PLCs). PLCs move professional development beyond the acquisition of new knowledge and skills (Vescio, Ross & Adams, 2007) to the reflection and improvement of classroom practice through collaboration among teachers (DuFour, 2004). PLCs reflect a shift away from traditional approaches of professional development which emphasize "knowledge for practice" to an approach grounded on the idea of "knowledge of practice" (Cochran-smith & Lytle, 1999). In this model, "teachers work together and engage in continual dialogue to examine their practice and student performance and to develop and implement more effective instructional practices" (Darling-Hammond & Richardson, 2009, p 3). While PLCs are a relatively new model, the notion

[18] This chapter is based on: Soares, F. & De Witte, K. Teachers' Collaboration and Student Achievement: An Instrumental Variables Approach.

of communities of practice has been in practice in many parts of Sub-Saharan Africa. In many countries in Sub-Saharan Africa teachers meet in school-based teacher groups or conferences **to** engage with their peers in local problem solving, share experiences and ideas, interpret the curriculum and reforms, and assist new teachers (Plessis & Muzaffar, 2010).

Although there is a broad consensus in the literature that the ultimate objective of teachers' collaboration and participation in PLCs is to increase student learning through improved teaching practices, there is a lack of rigorous empirical evidence investigating the effects of teachers' participation in collaborative activities on student learning, particularly in the context of developing countries. This paper contributes to the incipient body of empirical evidence on the impact of teachers' collaboration on student achievement in Sub-Saharan Africa, using Equatorial Guinea as a case study. This country is particularly interesting as the nation's oil wealth and political power is concentrated among the ruling elite, leading to high poverty levels, weak social indicators (IMF, 2009) and poor education outcomes (World Bank, 2017). The countries' authoritarian regime severely restricts freedom of association and assembly (Freedom House, 2016), which could affect teachers' level of comfort to work in collaborative structures. This paper uses a unique dataset that contains detailed and reliable information on student learning and teacher's characteristics. While similar data is easier to access in Western countries (Dunne et al., 2000; Louis & Marks 1998; Strahan, 2003; Supovitz, 2002), this is the first paper to analyze the effect of teachers' collaboration in developing countries in general, and Sub-Saharan Africa in particular.

This paper examines *to what extent does teachers' level of collaboration on pedagogical topics affect student achievement*? In addition, we explore the heterogeneous effects across different population groups. Given the deep ethnic divisions in Equatorial Guinea, we explore if teachers' likelihood to collaborate differ based on teachers' ethnic background. Currently, there is a lack of rigorous empirical evidence linking teachers' collaboration (or participation in more formal structures, such as PLCs) on teaching practices and student achievement. To the best of our knowledge two main systematic reviews have been conducted on the impact of participation on a PLCs on teaching practices and student achievement. Vescio, Ross & Adams (2008) reviewed studies assessing the impact of PLCs on teaching practices and student achievement. The authors found that all 8 research articles reviewed supported the idea that teachers' participation in PLCs can have a positive impact on student learning. One common feature among the 8 studies that facilitated success was teachers' collaborative efforts focused on the learning needs of the students. Lomos, Hofman and Roel (2011) explored the effects of PLCs on student achievement in secondary schools through a meta-analysis. The authors analyze five quantitative empirical studies that fall within their inclusion criteria. Results showed that effects of PLCs on student achievement are diverse in magnitude but positive. The systematic reviews emphasize the need for longitudinal observation studies

(both quantitative and qualitative) as well as experimental research that document changes in student learning as teachers work in PLCs (Vescio, Ross & Adams, 2007). Most of the evidence available is based on cross-sectional data, such that experimental (or quasi-experimental) and/or long-term designs are near absent in the literature. No studies from developing countries were included in the reviews.

From an empirical perspective, determining the impact of teachers' collaboration on student achievement is challenging because of omitted variable bias and the resulting endogeneity issues. A traditional Ordinary Least Square (OLS) estimate is not suitable for estimating the impact of teachers' level of collaboration on student learning. The OLS is based on the assumption that the model error term is unrelated to the regressors (Cameron & Trivedi, 2010). Endogeneity arises when this assumption is violated. Given that teachers with distinct levels of collaboration were not assigned randomly and exogenously to students, we expect the OLS regressor to be a biased estimate of the causal effect of teachers' collaboration on student achievement. Also, because teachers who exhibit high levels of collaboration may be different in unobservable ways from those who do not collaborate frequently, the relationship between teachers' collaboration and student achievement detected through an OLS estimation may be capturing unobservable influences on student learning that are omitted as predictors. To address these issues, we use an instrumental variable (IV) approach. In this paper, we use exogenous variation in number of teachers per school in the primary grades as an instrumental variable for teachers' level of collaboration and use this variation to estimate the impact of teachers' collaboration on student achievement. The IV method addresses issues of endogeneity and omitted variable bias, making it possible to investigate a causal relationship, rather than simply a correlation.

The remainder of the paper is structured as follows. In section 2, we present the major findings of the literature on teachers' participation in collaborative activities and student achievement. In section 3 the Equatorial Guinea setting is briefly described, while section 4 describes the data. Section 5 discusses the estimation strategy and Section 6 contains the empirical results. The final Section concludes with a discussion of main findings and policy implications.

4.2. Literature Review

Teachers traditionally work in isolation. Schools usually provide teachers with little opportunity for collaboration and ask them to make changes in their instruction in isolation and without any support (Hawley & Valli, 2000). This is reinforced as teachers often prefer to be left alone rather than to engage with their colleagues (Dufour, 2004) and tend to carry out their work autonomously, without assistance or input from others. Overall, schools do not provide sufficient opportunity and encouragement for teachers to work together, learn from each other, and improve their expertise as a community (Fullan & Hargreaves,

1996). The losses associated with the lack of collaboration among teachers have been recognized in the work of early educators, such as Dewey (1970): "*successes of [excellent teachers] tend to be born and die with them: beneficial consequences extend only to those pupils who have personal contact with the gifted teachers. No one can measure the waste and loss that have come from the fact that the contributions of such men and women in the past have been thus confined*." (p. 10). Lortie (1975) had also emphasized that isolation is detrimental to teachers as it limits access to new ideas and best practices.

Since the early work of Dewey (1970) and Lortie (1975), the literature has been converging into a consensus on the benefits of teachers' collaboration: it can potentially lead to improved teaching practices and increased student learning (Dufour, 2001, 2004; Fullan & Hargreaves 1991; Hawley & Valli, 2000; Little, 1990, 1999; Hord, 1997, 2004; Rosenholtz, 1989). This body of work reflects the view that quality teaching is not an individual accomplishment, but rather the results of collaborative processes that empower teachers to improve student learning beyond what any of them can achieve alone (Carroll, 2009). It recognizes that collaboration and reflection among a group of teachers exposes them to new ideas and practices and can work to improve their pedagogy through a process of critical reflection.

The literature has also advanced in the conceptualization of teachers' collaborative interactions and the essential characteristics of Professional Learning Communities (PLCs). Although there is no broad consensus in the literature, most definitions of PLCs seem to agree that it involves "a group of people sharing and critically interrogating their practice in an ongoing, reflective, collaborative, inclusive, learning-oriented, growth-promoting way (Stoll et al., 2006, p. 223). PLCs are based on the assumption that knowledge is situated in the daily experiences of teachers and best understood through critical reflection with others who share the same experience (Buysse, Sparkman & Wesley, 2003; Vescio, Ross & Adams, 2007). A key element of PLCs highlighted by Little (1990) is what she calls *joint work:* the thoughtful and explicit examination of teachers practices and their consequences. Joint work leads to learning and change as teachers describe, discuss, and adjust their practices according to a collectively held standard of teaching quality (Little, 2003).

Although there is a broad consensus in the literature that teachers' participation in a PLC can lead to increased student learning through improved teaching practices, the channels through which this change can happen are not clear. In fact, there is a risk that PLCs will reinforce and conserve existing traditional substandard practice rather than change it (Stoll et al., 2006). This point might be particularly relevant for Sub-Saharan Africa countries, where pedagogical practices may be resistant to change due to a discrepancy between what is required for teachers to teach under learner-centered approaches and the preparation they receive in initial teacher education, which is highly teacher-centered (Altinyelken, 2011; Akyeampong et al., 2013).

4.3. Context

Equatorial Guinea is a small Spanish-speaking country in Sub-Saharan Africa, identified by Freedom House as one of the nine least free countries in the world (Genocide Watch, 2012). The country is made up of a mainland territory called Rio Muni, and five islands including Bioko, where the capital Malabo is located. Since oil-discovery in the mid-1990s the country has become one of sub-Sahara's biggest oil producers, but a large proportion of its population still lives in poverty. There is deep ethnic division in Equatorial Guinea, and also clan division within ethnicities (Genocide Watch, 2012). According to the World Factbook (2017) the country's population is divided as following: Fang 85.7%, Bubi 6.5%, Mdowe 3.6%, Annobon 1.6%, Bujeba 1.1%, other 1.4%. The vast majority of the population and the ruling elite belong to the Fang group, which originates from Rio Muni. Within this group there are also clans. Nepotism and kleptocracy among the president's family and the Esangui clan of the Fang group has led to a concentration of the nation's oil wealth and political power (Genocide Watch, 2012). The minority groups have been historically excluded and discriminated against. The Bubi people, for example, are indigenous to Bioko Island and have been subject to systematic discrimination and persecution by the current government.

Equatorial Guinea' education system is characterized by low student learning outcomes and inefficiencies – a large numbers of learners in the early primary grades lack foundational reading and mathematics skills, while primary level drop-out and repetition rates are high (World Bank, 2017). Several factors contribute to the poor quality of education, including low teachers' qualification, lack of learning materials and poor physical conditions. While teacher training is improving, the rates of qualified teachers remain low and there are no mechanisms in place to hire or promote teachers based on their performance (World Bank, 2017). Teachers tend to depend heavily on dictation and copying and make limited use of problem solving and dialogues initiated by the students (World Bank, 2017).

4.4. Data

We use the 2015 Equatorial Guinea Primary Education Survey (EG-PES) database, developed by the Program for National Educational Development (PRODEGE) of Equatorial Guinea. PRODEGE is a program implemented by the Non-Government Organization Family Health International 360 (FHI360) in partnership with the Ministry of Education and Science (MoES) of Equatorial Guinea. The main objective of the survey is to collect information on teachers' classroom instructional practices and students' competences on reading and mathematics. The target population of the survey is restricted to students and teachers in first, third and sixth grades of primary education. These grades are chosen as they represent the beginning, middle and end of the primary education cycle in the country.

The EG-PES adopts a two-stage stratified random sampling by cluster strategy to select students and teachers from the target population. The sample is probabilistic (randomly selected), stratified by cluster, and selected in stages. The primary sampling unit (PSU) is the school (cluster) and the secondary sampling unit is the student or teacher. To ensure that schools are proportionately represented in the sample, the frame is stratified by the following variables: grade cycle offered; province; status of province capital for the two largest provinces.

EG-PES final sample consists of 66 schools: 33 schools that offer first and third grades (first cycle of primary education) and 33 schools that offer first, third and sixth grades (first and second cycle of primary education). Out of the 66 schools in the sample, three were no long in operation. The other 63 schools accepted to participate in the survey during the fieldwork. Within each school, the EG-PES collects data on 8 students per grade, totaling 480 grade1 students, 480 grade 3 students and 240 grade 6 students, selected randomly[19]. Given the length of the assessment, in some schools the data collection team was not able to collect data on all 8 students for each grade. No student refusal was recorded.

Within each school, each randomly selected student is invited to take a learning assessment test in mathematics and language, administered orally and on a one-on-one basis[20]. The learning assessment is followed by a background questionnaire, which asks questions on students' demographic and socio-economic characteristics. A structured interview questionnaire is applied to teachers of the selected grades and classrooms and asks questions related to teachers' socio-economic and demographic background, and participation in professional development activities. In some schools data collectors did not have enough time to apply the learning assessment to all 8 students selected in each grade within the number of days assigned to data collection for the school. As a consequence, although the targeted initial sample was 960 grades one and three students, only 914 students completed the Language assessment and 892 completed the Mathematics assessment. Since all 8 students per school were selected randomly and the order according to which they took the assessment was also random, we expect attrition to be random in this case.

The discussion provided under chapters two and three of this thesis show that the EG-PES complied with good standards highlighted in the survey methodology literature and identified and attempted to minimize potential survey errors where possible. As discussed in chapter 2, because the EG-PES followed best practices in sampling design it is fair to assume that the final sample represents the underlying population and that the empirical analysis conducted under the present chapter can be generalized to the population of

[19] A Kish selection table was developed as an instrument to assist field staff in randomly selecting classrooms and students to be part of the assessment.
[20] Two measures were taken to ensure inter-rater reliability: (i) during the training, enumerators worked in pairs and reviewed their coding after practice in the schools; (ii) at the end of the training, enumerators had to take a practical exam, and only those who scored 70% or above were selected for the survey.

grades 1, 3 and 6 students in Equatorial Guinea. A discussion on the development and psychometric properties of the EG-PES instruments in chapter three allowed us to provide evidence for the reliability and validity of the instruments and measures used.

Given the small sample size for sixth grade, this analysis focuses on first and third grades of primary education. The main variables of interest are defined as follows. *Student learning* is defined by two variables: the total student score in the mathematics assessment and the total score in the language assessment. Both variables are standardized with a mean of 0 and a standard deviation of 1. *Teachers' collaboration* is a binary variable that indicates high collaboration among teachers and takes the value of 1 if the teacher meets with others to discuss pedagogical topics once or more per month and 0 otherwise (i.e., teacher meets with other teachers once every other month or less frequent than that)[21]. If the teacher never meets with other teachers, the variable also takes the value of 0. As there is no theoretical or empirical recommendation in the literature as to how often teachers should collaborate on pedagogical topics to improve instructional practices and student learning, for the purpose of this research we define high collaboration as collaboration that happens at least once per month. A continuous variable for frequency of collaboration is not used as we are mostly interested in the impact of higher levels of teacher collaboration on student achievement. Since most schools do not keep records or logs of teacher meetings, this information is self-reported and is collected by asking teachers how often they meet to collaborate and work together on pedagogical topics, such as lesson planning and instructional practices. To avoid social desirability bias, before the start of the survey interviewers disclaimed to respondents that the information provided would be strictly confidential and protected by codes. Interviewers also made clear to respondents that he or she would not receive any personal benefit as a result of the survey. We believe teachers did not have any incentives or unobserved motives to over report their collaboration practices, since the country does not provide any incentive to teachers who collaborate more frequently. Also, to the best of our knowledge, at the time of the survey, collaboration was not a practice incentivized or promoted by local or national policies. *Teachers' Fang ethnic background* refers to teachers' affiliation with the Fang ethnic group. This is a binary variable that takes the value of 1 if teacher is Fang, and 0 otherwise. Given the ethnic tensions and divisions in the country, asking teachers and students directly about their group affiliation was considered a sensitive question that could make respondents uncomfortable, potentially making their responses biased. To avoid this issue, instead the survey asked students and teachers to report the language they usually speak at home as a proxy for ethnic group affiliation. This proxy is based on the assumption

[21] Due to data constraints more flexible models that account for number of meetings per year are not possible to estimate.

that the concepts of ethnicity and language are inter-related and ethnicity is sometimes used as an umbrella term, with language as a key identity marker for an ethnic group (Education Equity Research Initiative, 2016). It is important to note, however, that there is a tendency observed in the country where families are switching from their ethnic language to Spanish as the main language spoken at home. For the teachers that reported speaking Spanish at home, we do not have information on their ethnic group affiliation: they may or may not have a Fang ethnic background.

Table 4.1 presents descriptive statistics on key variables of interest. On average, 48.5% of teachers interviewed reported speaking Fang at home. Also, approximately 48% of the teachers reported collaborating with other school teachers on pedagogical topics once per month or more often. The average total number of primary grade teachers per school is 4 teachers. The average total number of Fang teachers per school is 1.5. The actual average total number of Fang teachers per school is likely higher than this, as the variable only captures teachers that report speaking Fang at home.

Table 4.1. Descriptive statistics on key variables at student, teacher and school level

Variable	Observations	Mean	St. Dev.	Min	Max
Language standardized score	914	0.000	1.000	-2.314	2.299
Mathematics standardized score	892	0.000	1.000	-2.334	2.333
Teacher speaks Fang at home	153	0.484	0.501	0	1
Teacher exhibit high collaboration	153	0.477	0.501	0	1
Total number of teachers grades 1-3 in school	63	4.032	4.142	1	18
Total Fang teachers grades 1-3 in school	63	1.524	1.255	0	6

4.5. Empirical Approach

Naïve OLS regression and choice of the instrument

Table 4.2 presents the results of the 'naive' OLS regressions, which do not account for the endogeneity issues. We regress teachers' high collaboration on standardized mathematics and language scores, controlling for student and teacher socio-demographic characteristics. Given the hierarchical structure of our data we employ cluster-robust standard errors to account for within cluster correlation. Specifically, we have cross-sectional data on students with clustering on schools. It is important to control for within-cluster error correlation as failure to do so can lead to misleadingly small standard errors.

We observe a positive, though insignificant, relationship between teachers' high level of collaboration and language scores. As noted earlier, the OLS estimates are probably biased due to endogeneity issues and mathematics and language scores are likely strongly influenced by observed and unobserved heterogeneity.

Table 4.2. Estimation of teachers' high collaboration on standardized mathematics and language scores by 'naïve' OLS regression

Variable	Language	Mathematics
High collaboration	0.00524	-0.0265
	(0.0666)	(0.0700)
Constant	-0.749***	0.498*
	(0.244)	(0.262)
Controls	YES	YES
Observations	914	892
R-squared	0.137	0.094

Notes: (i) robust standard errors are adjusted for clustering by school. (ii) *** indicates statistical significance at 1%, ** at 5% and * at 10%. Full regression outcomes are provided in Appendix.

Instrumental variables

To address the endogeneity issues, we use number of teachers in primary grades one through three within the school as an instrumental variable for teachers' collaboration on pedagogical topics. In contrast to other education systems, the number of teachers is not a choice of the school in Equatorial Guinea and can be considered as exogenous: the central government assigns teachers to both public and private schools in a somewhat arbitrary manner without undergoing a consultation process with the local schools or the teachers to be assigned.

The instrumental variable strategy is valid as long as the number of teachers only affect student learning via teachers' collaboration, and not through other channels. This means that number of teachers should be correlated to teachers' collaboration, but not directly to student learning. In developing countries, many teachers, particularly those teaching in rural areas or in small multi-grade schools, are often isolated and disconnected from their peers, which hinders their opportunities to collaborate with one another. Teachers based in schools with smaller number of teachers, have less opportunities to collaborate with one another. Not surprisingly, we find that number of teachers in primary grades one through three within the school is correlated with teachers' collaboration (corr = .054; p = .10).

There are potential concerns with the choice of instrument related to the exclusion restriction. One possibility might be that a high number of teachers in primary grades positively affect student learning by making more teachers available to students and possibly reducing student-pupil ratio. Or, alternatively, that a high number of teachers in primary grades negatively affect student learning by increasing bureaucratic processes teachers need to go through. However, these concerns do not appear to be a serious threat to our strategy. First, we have reason to believe that the hiring of teachers is not correlated to total enrollment and student-pupil ratio. There are some indications the government may at times use enrollment data to inform the assignment of teachers to schools, but this is not done through a systematic process where an explicit rule of having one more student enrolled would prompt another teacher hiring. Second, there is no clear link established in the literature between school size (as measured by total enrollment or total number of teachers) and student achievement. Glewwe et al. (2011) examined studies from developing countries to investigate which specific school and teacher characteristics, if any, have positive impacts on learning. They found that the evidence on the impact of school size on learning is inconclusive: that the overall size of the school has no clear tendency, and it is not clear a priori what the sign of the effect should be. In summary, we believe that number of teachers in primary grades as an instrument fulfill the two critical conditions for an unbiased estimate: exogeneity and relevance.

Table 4.3. Estimation of number of teachers in primary grades one through three on mathematics and language scores by OLS regression

Variable	Language	Mathematics
Number of teachers in primary grades 1-3	0.0138	-0.00563
	(0.0123)	(0.0117)
Constant	-0.575*	0.735**
	(0.314)	(0.353)
Controls	YES	YES
Observations	914	892
R-squared	0.151	0.114

Notes: robust standard errors are adjusted for clustering by school. *** indicates statistical significance at 1%, ** at 5% and * at 10%.

We use a two-stage least square estimation (2SLS) procedure to estimate the IV regressions. To account for the binary nature of the endogenous regressor used in this analysis (teachers' collaboration), we apply

a latent-variable model in the first-stage of the 2SLS procedure (Cameron & Trivedi, 2010). The model introduces an unobserved latent variable, X^*, that determines whether $X' = 1$ or 0. The model becomes:

$$\textit{First-stage: } X_i^* = a_0 + \beta_1 Z_i + \beta_\mu \Omega_i + \beta_{\mu+1} \Omega_i + \square\ \beta_{\mu+\rho} \Omega_i + e_i \quad (1)$$

$$\textit{Second-stage: } Y_i = \langle_0 + ®_1 Xi' + ®_2 W_i + \square + ®_{1+\rho} W_i + \sum_i \quad (2)$$

$X'_i = 1$ *if* $X_i^* > 0$

$X'_i = 0$ otherwise

where Y_i is the outcome variable for observation *i*; X_i is the ith observation of the binary endogenous explanatory variable; X_i' is the fitted value of X_i; W_i …. W_i are the ith observation of each of covariates; and e_i *and* ε_i are error terms. We employ cluster-robust standard error to account for within cluster correlation.

4.6. Results

4.6.1. Full sample analysis: first stage results

Table 4 presents the results of the first-stage regressions, which test the relevance of the instrument[22]. We report results for first-stage equations for both outcomes analyzed in the second stage – scores in the language assessment and scores in the mathematics assessment. We would expect the first stage results to be the same for both outcomes analyzed in the second stage, but because the sample size for students that took the language and the mathematics assessments are different, the first-stage results differ slightly. Overall, we observe a positive significant relationship between the number of teachers in primary grades one through three and teacher collaboration. This indicates that an increase by one teacher in the total number of teachers in primary grades one through three increases the probability of teachers' collaboration by 3.5% under the language achievement analysis and by 2.9% under the mathematics achievement analysis. As a rule of thumb, the t-value for the instrument should be equal to or larger than 3.2. Results in Table 4.4 show a t-statistic of 3.14 for Language, which is just slightly below the 3.2 threshold, and 2.47 for Mathematics. These estimates show that the instrument comply with the relevance assumption for Language, but not for Mathematics. This evidence is in line with various tests on the reliability and

[22] Various tests on the reliability and efficiency of the instrument are provided in Appendix II. These statistics suggest that teacher collaboration is an endogenous regressor when analyzing student performance in language and there is a need to instrument it; the model is not under identified; and there is no problem of a weak instrument as all F statistics are above the lowest critical value of 5.53.

efficiency of the instrument are provided in Appendix 4.B, which suggest a need to instrument teacher collaboration when analyzing student performance in Language. Full regression results are provided in Appendix 4.C.

Table 4.4. First-stage regression results: relationship between high levels of collaboration and total number of teachers in primary grades

	Language	Mathematics
Number of teachers in grades 1-3	0.0349***	0.0289**
	(0.0111)	(0.0117)
Constant	-0.427	-0.412
	(0.333)	(0.340)
Controls	YES	YES
Observations	914	892
R-squared		

Notes: (i) robust standard errors are adjusted for clustering by school; (ii) controls are included for student and teacher background characteristics; (iii)*** indicates statistical significance at 1%, ** at 5% and * at 10%. Full regression results are provided in Appendix 4.C.

4.6.2. Full sample analysis: second stage results

The second stage results are presented in Table 4.5 using regular 2SLS estimation and applying a latent-variable model in the first-stage of the 2SLS procedure. Full regression tables are presented in Appendix 4.C. The results show that there is a positive and significant effect of teachers' collaboration around pedagogical topics on the performance of students in the language assessment. The results show that, on average, teachers that exhibit a high level of collaboration cause an increase of approximately 0.5 standard deviations in the score in the language assessment, which is regarded as a large effect by the education literature, even in developing countries. Overall, an increase of less than 0.1 standard deviation in student learning is typically considered to be a small effect, while an increase of more than 0.3 standard deviation is considered a large effect (Masino & Niño-Zarazúa, 2016). One can expect that effect sizes are larger in developing countries as educational innovations are less prone to ceiling effects which are typically observed in developed countries. On the other hand, the large effect observed might be related to the binary nature of the variable collaboration, which measures two extremes of collaboration – high versus low levels of collaboration. As we can see in table 4.5, the results do not show a significant effect of teachers' collaboration in the mathematics assessment. These results are not consistent with the literature on teacher-value added by subject area, which have found a larger influence of teacher quality on math. Hanushek and

Rivkin (2010) review estimates of the within-school variance in teacher quality and find a lower average for reading (0.11 SD) when compared to math (0.15 SD). Unfortunately, we lack data on the nature and focus of collaboration between teachers, which may cover topics on mathematics instruction, reading instruction or other. Given the Ministry of Education's focus on improving student reading outcomes since 2006 and the development and distribution of materials to improve reading instruction, it is likely that teachers' collaborative activities placed a higher focus on language. This would explain why we find a significant effect of teachers' collaboration in language, but not in mathematics.

We observe that the results of the OLS estimation (see section 5.2) and the IV estimation differ considerably, in particularly for the language assessment. The influence of teachers' collaboration in the language scores is not significant in the OLS estimation, while in the IV estimation this impact is significantly positive. We also observe that the influence, in comparison to the OLS estimation, is much larger. We can conclude that our concerns regarding endogeneity issues for language as our main variable of interest are justified. For the mathematics assessment, we do not observe a large difference between the OLS and IV estimators; both estimators remain negative and not significant, although the influence observed under the IV estimation is much larger. The lack of major differences between the OLS and IV estimators for the mathematics assessment is expected and consistent with the tests in the reliability of the instrument (Appendix 4.B.), which failed to reject the null hypothesis that the regressor teacher collaboration is exogenous for mathematics as the main outcome of interest.

Three covariates significantly correlate to scores in the language assessment**.** We observe a negative significant relationship between student age and student achievement. This finding is consistent with research from the Programme for the Analysis of Education Systems (PASEC) in other Sub-Saharan African countries, which indicate that over-aged students have lower achievement when compared to regular-age students (Malpel, 2016). We also observe a positive significant relationship between student wealth index and student score in the language assessment. This is consistent with the literature indicating that socioeconomically disadvantaged students tend to perform worse on standardized measures of academic achievement compared to their wealthier peers (Noel & de Broucker, 2001; Perry & McConney, 2010; PISA, 2004; Sirin, 2005). We also find a significant negative relationship between teachers that speak Spanish at home and student achievement. Teachers who speak Spanish at home are likely to be more proficient and comfortable with the language. Since Spanish is the main language of instruction, we would expect a direct impact of teachers' proficiency on student language scores.

Table 4.5. Two-stage least square results: relationship between teachers' collaboration and student score in language and mathematics assessment

	Language	Mathematics
Teacher exhibit high collaboration	0.543*	-0.137
	(0.320)	(0.564)
Constant	-0.967***	0.541
	(0.274)	(0.330)
Controls	YES	YES
Observations	914	892

Notes: (i) teachers' collaboration is instrumented by the number of teachers per school in grades 1-3; (ii) robust standard errors are adjusted for clustering by school; (iii) controls are included for student and teacher background characteristics; (iv)*** indicates statistical significance at 1%, ** at 5% and * at 10%. Full regression results are provided in Appendix 4.C.

4.6.3. Heterogeneous effects by ethnic groups

We conduct a heterogeneity analysis to identify the impact of teachers' collaboration on student learning among Fang teachers only. Given the deep ethnic divisions in Equatorial Guinea, we expect Fang teachers, which represent the majority and historically dominant tribe in the country, to be more likely to collaborate with one another, following historical patterns of exclusion and discrimination against minority groups. We restrict the sample to Fang teachers only and we use the total number of Fang teachers in primary grades within the school as the instrument. Within this restricted sample, we also estimate the same model described above for each outcome of interest (score in the mathematics assessment and score in the language assessment). The naïve OLS and the first stage tests are provided in Appendix 4.D. In Appendix 4.A. we also examine differences in observable characteristics of Fang teachers compared to non-Fang teachers. The results indicate no differences between the two groups across most characteristics, with the exception of gender and teaching certificate. Surprisingly, a significantly higher proportion of non-Fang teachers report having a teaching certificate.

4.6.4. Sample restricted to Fang teachers: first stage results

Table 4.6 presents the results of the first-stage regression using the sample restricted to Fang teachers only. The results under both outcomes of interest – scores in the language assessment and scores in the mathematics assessment, indicate a positive significant relationship between the number of Fang teachers in primary grades one through three and collaboration among Fang teachers. Full regression results are provided in Appendix 4.F. When comparing the first-stage results for the sample restricted to Fang teachers with the first-stage results for the full sample, we observe that the coefficients for the influence of number

of teachers in primary grades on the probability of teachers exhibiting high levels of collaboration is much larger in magnitude among the restricted sample of Fang teachers. As the majority and dominant ethnic group, Fangs have historically excluded and discriminated against other minority groups, which might explain the higher collaboration levels between Fang teachers.

Table 4.6. First-stage regression results: relationship between high levels of collaboration and total and total number of fang teachers in primary grades

	Language	Mathematics
Number of fang teachers in grades 1-3	0.154***	0.176***
	(0.0554)	(0.0588)
Constant	0.615	0.538
	(0.482)	(0.495)
Control	YES	YES
Observations	443	430
R-squared		

Notes: (i) robust standard errors are adjusted for clustering by school; (ii) controls are included for student and teacher background characteristics; (iii)*** indicates statistical significance at 1%, ** at 5% and * at 10%. Full regression results are provided in Appendix 4.F.

4.6.5. Sample restricted to Fang teachers: second stage results

Table 4.7 reports the results for the second stage regressions restricted to the Fang teachers' sample. Teachers' level of collaboration around pedagogical topics has a positive impact on language achievement. Fang teachers that exhibit high levels of collaboration lead to a 0.583 standard deviation increase in student score in the language assessment. We also observe a positive impact of Fang teachers' collaboration on student achievement in mathematics. The impact is marginally significant (p=0.115). The impact of Fang teachers' collaboration in the language assessment is similar in magnitude to the results for the entire sample (Table 5) and is considered large by standards of the education literature, as previously discussed. However, we observe a much larger impact in the mathematics assessment when restricting the sample to Fang teachers only. This may be partially explained by the fact that teacher collaboration is an endogenous regressor for the sample restricted to Fang teachers, but not for the entire sample. As shown in Table 4.D.1 of Appendix 4.D, tests of reliability of the instrument for the Fang teacher sample reject the null hypothesis that the regressor (i.e. teacher collaboration) is exogenous for both language and mathematics score as our main outcome of interest. Additionally, as the probability of teachers exhibiting high levels of collaboration is stronger among Fang teachers (as per results of Table 4.6), it is possible that the nature of their

collaboration is also different. Higher levels of collaboration might be associated with more in-depth discussions and a focus that expands beyond language instruction to also include mathematics.

As indicated by the full regression results in Appendix 4.F, one covariate significantly correlates to scores in the language and mathematics assessments. As we found for our analysis using the full sample, there is significant positive relationship between student wealth index and student scores in the both assessments.

Table 4.7. Two-stage least square results: relationship between teachers' collaboration and student score in language and mathematics assessment for Fang teachers only

	Language	Mathematics
Teacher exhibit high collaboration	0.583*	0.529
	(0.342)	(0.335)
Constant	-1.729***	-0.190
	(0.429)	(0.438)
Controls	YES	YES
Observations	443	430

Notes: (i) teachers' collaboration is instrumented by the number of Fang teachers per school in grades 1-3; (ii) robust standard errors are adjusted for clustering by school; (iii) controls are included for student and teacher background characteristics; (iv)*** indicates statistical significance at 1%, ** at 5% and * at 10%. Full regression results are provided in Appendix 4.F.

4.7. Discussion of the Results and Policy Implications

Collaboration among teachers has been increasingly promoted in developing countries. The idea that teachers' collaboration can lead to an increase in student learning through improved teaching practices, has gained traction in the literature and in the practice of international development organizations and governments. Nonetheless, there is a lack of rigorous empirical evidence investigating the effects of teachers' participation in collaborative activities on student learning, particularly in the context of developing countries.

This paper contributes to the literature by examining the relationship between teachers' level of collaboration and student achievement in primary grades in Equatorial Guinea. We use an instrumental variable (IV) approach to address issues of endogeneity and omitted variable bias. By using the exogenous variation in number of teachers per school in the primary grades as an instrumental variable for teachers' level of collaboration**,** we explore the causal relationship between teachers' collaboration and student

achievement. In addition, we also conducted a heterogeneity analysis to identify the impact of teachers' collaboration among Fang teachers only (majority and dominant ethnic group) on student achievement.

Our results show that teachers' high level of collaboration around pedagogical topics has a positive influence on student achievement in language, but not in mathematics. Our results are stronger when we restrict our sample to Fang teachers only. Given that there is a lower proportion of Fang teachers who receive a teaching certificate compared to non-Fang teachers, it is unlikely that these results are being driven by the training background of Fang teachers. When comparing the coefficients for the first-stage results of both analysis, we observe that the influence of number of teachers in primary grades on the probability of teachers exhibiting high levels of collaboration is higher among the restricted sample of Fang teachers. These results indicate that teachers' background may influence their likelihood to collaborate, as teachers may be more willing to collaborate with other teachers that share the same background characteristics. This is in line with previous literature highlighting that teacher professional communities are more likely to occur in homogeneous rather than heterogeneous groups (Bryk et. al., 1999). Bryk et al. (1999) explain that "when all members of the group share the same assumptions, habits, and values, it is reasonably easy to engender the trust and stable expectations for behavior that support community" (p.758). This may be especially true in the context of developing countries with marked ethnic group divisions, such as Equatorial Guinea.

This research has important policy implications. Our findings support the establishment and growth of formal collaborative structures and initiatives within schools and among schools, such as PLCs and communities of practice, in Sub-Saharan Africa countries. Our analysis also indicated stronger collaboration among teachers from the same ethnic-majority group and stronger impact on student learning. This finding suggests that the composition of collaborative structures by teachers' background may have a different impact on student learning. Overall, there is a need to better understand the specific features of these potential collaborative structures for increased collaboration among teachers of different backgrounds and an optimal impact on student learning.

We acknowledge an important limitation of this study. The data used in the analysis does not contain information on the nature of the collaborative process that takes place among the teachers and the specific types of discussion and pedagogy topics addressed during their meetings. Collaboration among teachers can vary greatly, ranging from superficial exchanges to in-depth reflection and discussion of teaching practices and analysis of student work. Without this information, we cannot know which specific interactions and dynamics of collaboration may lead to teacher learning, innovations in teaching practices and improved student learning. Further research investigating the channels through which teacher collaboration can lead to student learning is warranted.

References

Altinyelken, H. K. (2010). Pedagogical renewal in sub- Saharan Africa: the case of Uganda. *Comparative Education*, *46*(2), 151–171.

Akyeampong, K., Lussier, K., Pryor, J., & Westbrook, J. (2013). Improving teaching and learning of basic maths and reading in Africa: Does teacher preparation count? *International Journal of Educational Development*, *33*(3), 272–282.

Bryk, A., Camburn, E., & Louis, K. S. (1999). Professional community in Chicago elementary schools: Facilitating factors and organizational consequences. *Educational administration quarterly*, *35*(5), 751-781.

Buysse, V., Sparkman, K. L., & Wesley, P. W. (2003). Communities of practice: Connecting what we know with what we do. *Exceptional children*, *69*(3), 263-277.

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, *50*(2), 317-372.

Cameron, A. C., & Trivedi, P. K. (2010). *Microeconometrics using stata* (Vol. 2). College Station, TX: Stata press.

Carroll, T. (2009). The next generation of learning teams. Phi Delta Kappan, 91(2), 8–13

Cochran-Smith, M., & Lytle, S. L. (1999). Chapter 8: Relationships of knowledge and practice: Teacher learning in communities. *Review of research in education*, *24*(1), 249-305.

Davidson, R., & MacKinnon, J. G. (1993). Estimation and inference in econometrics. *OUP Catalogue*.

Darling-Hammond, L., & Richardson, N. (2009). Research review/teacher learning: What matters. *Educational leadership*, *66*(5), 46-53.

Dewey, J. (1970). *The sources of a science of education*. New York: Liveright.

Du Plessis, J., & Muzaffar, I. (2010). Professional Learning Communities in the Teachers' College: A Resource for Teacher Educators. *EQUIP1*.

DuFour, R. (2004). What is a" professional learning community"?. *Educational leadership*, *61*(8), 6-11.

Dunne, F., Nave, B., & Lewis, A. (2000). Critical friends groups: Teachers helping teachers to improve student learning. *Phi Delta Kappan*, *28*(4), 31-37.

Education Equity Research Initiative (2016). Measuring Equity in Education: Review of the Global and

Programmatic Data Landscape. Retrieved from: https://static1.squarespace.com/static/57aa9cce6b8f5b8163fdc9a3/t/582bdaebe3df28e79513dc46/1479269100751/Equity+Initiative+Measurement+and+Metrics+Landscape+Review+2016.pdf

Freedom House. (2017). Freedom in the World 2017. Populists and autocrats: The dual threat to global democracy. *Freedom House*.

Freedom House (2016). Equatorial Guinea: Freedom in the World 2016. Retrieved from: https://freedomhouse.org/report/freedom-world/2016/equatorial-guinea

Fullan, M., & Hargreaves, A. (1996). *What's Worth Fighting for in Your School? Revised Edition*. Teachers College Press, 1234 Amsterdam Avenue, New York, NY 10027.

Genocide Watch (2012). Updated country profile of Equatorial Guinea. Retrieved June 30, 2019, from http://genocidewatch.net/2013/03/19/country-profile-equatorial-guinea/

Glewwe, P. W., Hanushek, E. A., Humpage, S. D., & Ravina, R. (2011). *School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010* (No. w17554). National Bureau of Economic Research.

Hanushek, E. A. (1995). Interpreting Recent Research on Schooling in Developing Countries. *The World Bank Research Observer*, *10*(2), 227–246.

Hawley, W. D., & Valli, L. (2000). Learner-centered professional development. *Phi Delta Kappa Center for Evaluation, Development, and Research*, *27*, 7-10.

Hord, S. M. (1997). Professional learning communities: Communities of continuous inquiry and improvement.

Hord, S. M. (Ed.). (2004). *Learning together, leading together: Changing schools through professional learning communities*. Teachers College Press.

Little, J. W. (1990). The persistence of privacy: Autonomy and initiative in teachers' professional relations. *Teachers college record*, *91*(4), 509-536.

Little, J. W. (2003). Inside teacher community: Representations of classroom practice. *Teachers college record*, *105*(6), 913-945.

Lortie, D. (1975). Schoolteacher: A sociological analysis. *University of Chicago*.

Louis, K. S., & Marks, H. M. (1998). Does professional community affect the classroom? Teachers' work and student experiences in restructuring schools. *American journal of education*, *106*(4), 532-575.

Masino, S., & Niño-Zarazúa, M. (2016). What works to improve the quality of student learning in developing countries?. *International Journal of Educational Development*, *48*, 53-65.

Noël, S., & de Broucker, P. (2001). Intergenerational inequities: A comparative analysis of the influence of parents' educational background on length of schooling and literacy skills. *In pursuit of equity in education: Using international indicators to compare equity policies*, 277-298.

Orr, D., Westbrook, J., Pryor, J., Durrani, N., Sebba, J., Adu-Yeboah, C. (2013). *What are the impacts and cost-effectiveness of strategies to improve performance of untrained and under-trained teachers in the classroom in developing countries?* London: EPPI-Centre, Social Science Research Centre, Institute of Education, University of London.

PASEC (2015). PASEC 2014: Education system Performance in Francophone Sub-Saharan Africa. Retrieved from: http://www.pasec.confemen.org/wp-content/uploads/2015/12/Rapport_Pasec2014_GB_webv2.pdf

Perry, L. B., & McConney, A. (2010). Does the SES of the school matter? An examination of socioeconomic status and student achievement using PISA 2003. *Teachers College Record*, *112*(4), 1137-1162.

Programme for International Student Assessment. (2004). *PISA Learning for Tomorrow's World: First Results from PISA 2003* (Vol. 659). Simon and Schuster.

Rosenholtz, S. J. (1989). *Teachers' workplace: The social organization of schools*. Addison-Wesley Longman Ltd.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of educational research*, *75*(3), 417-453.

Stock, J. H., & Yogo, M.. (2005). *Testing for Weak Instruments in Linear IV Regression.* In D. W. K. Andrews and J. H. Stock (Ed.), Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg (pp. 80-108). New York, NY: Cambridge University Press.

Stoll, L., Bolam, R., McMahon, A., Wallace, M., & Thomas, S. (2006). Professional learning communities: A review of the literature. *Journal of educational change*, *7*(4), 221-258.

Strahan, D. (2003). Promoting a collaborative professional culture in three elementary schools that have beaten the odds. *The Elementary School Journal*, *104*(2), 127-146.

Supovitz, J. A. (2002). Developing communities of instructional practice. *Teachers College Record*, *104*(8),

1591-1626.

*The World Factbook (*2017). Washington, DC: Central Intelligence Agency, 2017.

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2010). *Reaching the marginalized* (EFA Global Monitoring Report 2010). Paris, France: Author.

World Bank (2017). Guinea Ecuatorial: diagnóstico del sector educación.

Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and teacher education*, *24*(1), 80-91.

Appendix 4.A. Descriptive statistics

Table 4.A.1. Descriptive statistics

Variable	Observations	Mean	St. Dev.	Min	Max
Student age	914	9.470	2.188	5	16
Student female	914	0.492	0.500	0	1
Student household size	914	6.995	2.884	2	27
Student overage	914	0.458	0.499	0	1
Student repeating grade	914	0.189	0.392	0	1
Student wealth index missing	914	0.0470	0.212	0	1
Student wealth index	914	2.010	0.714	0.0111	3.200
Student speaks Spanish at home	914	0.568	0.496	0	1
Student has access to sewerage	914	0.556	0.497	0	1
Student favorite subject Language	914	0.306	0.461	0	1
Language standardized score	914	-0.000671	1.000	-2.314	2.299
Mathematics standardized score	892	-0.000651	1.000	-2.334	2.333
Teacher age	153	38.33	11.58	19	66
Teacher female	153	0.281	0.451	0	1
Teacher years of experience	153	11.32	8.633	1	46
Teacher received training	153	0.451	0.499	0	1
Teacher has teaching certificate	153	0.124	0.331	0	1
Teacher speaks Fang at home	153	0.484	0.501	0	1
Teacher speaks Spanish at home	153	0.366	0.483	0	1
Teacher is hired by the government	153	0.667	0.473	0	1
Teacher exhibit high collaboration	153	0.477	0.501	0	1
Total number of teachers grades 1-3	63	4.032	4.142	0	18
Total Fang teachers grades 1-3	63	1.524	1.255	0	6

Table 4.A.2. Distribution of frequency with which teacher meets with other teachers to collaborate on pedagogical topics

	Observations	Frequency	Cumulative Freq.
Never	294	32.27	32.27
Once a year	39	4.28	36.55
Once every semester	116	12.73	49.29
Once every two months	20	2.2	51.48
Once a month	125	13.72	65.2
Twice a month	28	3.07	68.28
Weekly	289	31.72	100

Table 4.A.3. Teachers' background characteristics by Fang status

	Fang		Non-Fang		
	Observations	Mean	Observations	Mean	Diff
Teacher age	74	39.45	79	37.29	-2.15
Teacher female	74	0.38	79	0.19	-0.19***
Teacher years of experience	74	11.78	79	10.89	-0.90
Teacher received training	74	0.47	79	0.43	0.04
Teacher has teaching certificate	74	0.05	79	0.19	0.14**
Teacher is hired by the government	74	0.68	79	0.66	0.02

*Note: Asterisks indicate statistical significance: * .10 ** .05 *** .001*

Appendix 4.B. Instrumental variable first stage tests for full sample

Table 4 summarizes the statistical tests to assess the reliability and the efficiency of the IV estimation. The Durbin-Wu-Hausman test tests the null hypothesis that the regressor (i.e. teacher collaboration) is exogenous. The test, suggested by Davidson and MacKinnon (1993), includes the residuals of each endogenous right-hand side variable, as a function of all exogenous variables, in a regression of the original model. Results from Table A.2 indicate that the null hypothesis is rejected for language score as the main outcome of interest but not for mathematics score. This indicates that when analyzing student performance in language, teacher collaboration is an endogenous regressor and there is a need to instrument it. For mathematics, we expect that there is only a little difference between OLS and IV estimators. Moreover, it suggests that the naïve OLS outcomes are not significantly biased.

The Kleibergen-Paap rk LM test examines the null hypothesis of underidentification. For both outcome variables of interest (scores in language and mathematics) we can reject the null hypothesis and conclude that the model is identified.

We use two tests of weak instruments: the Cragg- Donald Wald F statistic and the Kleibergen-Paap Wals rk F statistics. Both tests aim to address the concern that the estimation bias of the IV estimator resulting from the use of weak instruments can be large (sometime even larger than the bias of OLS estimators) (Cameron & Trivedi, 2010). We use Stock-Yogo critical values tabulated for the Cragg-Donald F statistic (Stock & Yogo, 2005). The first-stage Cragg- Donald Wald F statistics is 6.477 for Mathematics and 8.441 for language, both above the critical value of 5.53 for a relative 25% bias toleration. The Kleibergen-Paap Wals rk F statistics for mathematics is above the 6.66 critical value (20% bias toleration) and for language the F statistic is above 8.96 (15% bias toleration). We conclude that there is no problem with weak instruments as all F statistics are above the lowest critical value of 5.53.

Tests of overidentifying restrictions using the Hansen J statistics are not performed as our model is just-identified. These tests can only be performed in overidentified models, i.e., when the number of instruments exceeds the number of endogenous regressors.

Table 4.B.1. Chapter 4 first stage tests for regressions using full sample

	Language	Mathematics
Endogeneity test		
Durbin- Wu-Hausman (chi squared) test	3.81126*	0.133971
Under-identification test		
Kleibergen-Paap rk LM statistics	8.487**	6.389**
Weak identification test		
Cragg- Donald Wald F statistic	8.441	6.477
Kleibergen-Paap Wals rk F statistics	9.769	7.203
Overidentification test		
Hansen J statistics		

(1) where ***, **, * denote significance at 1%, 5% and 10%-level

(2) Stock-Yogo weak ID test critical values: 10% = 16.38; 15% = 8.96 ; 20% = 6.66; 25% = 5.53

Appendix 4.C. Results Tables for Full Sample

Table 4.C.1. Estimation of teachers' high collaboration on mathematics and language scores by 'naïve' OLS regression

Variable	Language	Mathematics
Teacher exhibit high collaboration	0.00524	-0.0265
	(0.0666)	(0.0700)
Grade	0.320***	-0.184***
	(0.0414)	(0.0435)
Student age	-0.0394*	-0.0102
	(0.0213)	(0.0220)
Student female	0.123*	-0.0559
	(0.0626)	(0.0653)
Student household size	0.0175*	0.00275
	(0.0104)	(0.0116)
Student wealth index missing	-0.217	-0.103
	(0.151)	(0.158)
Student wealth index	0.198***	0.136***
	(0.0482)	(0.0502)
Student favorite subject Language	-0.0820	-0.185***
	(0.0680)	(0.0709)
Teacher age	-0.000905	-0.00293
	(0.00384)	(0.00421)
Teacher female	0.00344	-0.0638
	(0.0708)	(0.0755)
Teacher years of experience	0.000312	0.00466
	(0.00528)	(0.00581)
Teacher has teaching certificate	0.172	0.0725
	(0.106)	(0.104)
Teacher speaks Spanish at home	0.0959	0.128*
	(0.0664)	(0.0690)
Teacher is hired by the government	-0.136*	-0.326***
	(0.0775)	(0.0831)
Constant	-0.749***	0.498*
	(0.244)	(0.262)
Observations	914	892
R-squared	0.137	0.094

Notes: (i) robust standard errors are adjusted for clustering by school. (ii) *** indicates statistical significance at 1%, ** at 5% and * at 10%.

Table 4.C.2. **First-stage regression results**: relationship between high levels of collaboration and total number of teachers in primary grades

	Language	Mathematics
Number of teachers in Grades 1 - 3	0.0349***	0.0289**
	(0.0111)	(0.0117)
Grade	-0.0459	-0.0514
	(0.0587)	(0.0598)
Student age	0.106***	0.104***
	(0.0281)	(0.0284)
Student female	0.00335	0.0159
	(0.0879)	(0.0890)
Student household size	-0.000399	0.00316
	(0.0150)	(0.0155)
Student wealth index missing	0.321	0.320
	(0.208)	(0.217)
Student wealth index	-0.276***	-0.260***
	(0.0671)	(0.0680)
Student favorite subject Language	0.0240	0.0429
	(0.0947)	(0.0964)
Teacher age	0.00935	0.00987
	(0.00604)	(0.00607)
Teacher female	0.0121	-0.00930
	(0.106)	(0.107)
Teacher years of experience	-0.00693	-0.00861
	(0.00825)	(0.00823)
Teacher has teaching certificate	0.609***	0.621***
	(0.146)	(0.146)
Teacher speaks Spanish at home	-0.478***	-0.493***
	(0.0920)	(0.0928)
Teacher is hired by the government	-0.530***	-0.546***
	(0.110)	(0.111)
Constant	-0.427	-0.412
	(0.333)	(0.340)
Observations	914	892

Notes: (i) robust standard errors are adjusted for clustering by school. (ii) *** indicates statistical significance at 1%, ** at 5% and * at 10%.

Table 4.C.3. **Two-stage least square results**: relationship between teachers' collaboration and student score in language and mathematics assessment

	Language	Mathematics
Teacher exhibit high collaboration	0.543*	-0.137
	(0.320)	(0.564)
Grade	0.325***	-0.185***
	(0.0436)	(0.0444)
Student age	-0.0590**	-0.00623
	(0.0235)	(0.0290)
Student female	0.122*	-0.0552
	(0.0648)	(0.0650)
Student household size	0.0176	0.00292
	(0.0112)	(0.0114)
Student wealth index missing	-0.283*	-0.0897
	(0.156)	(0.170)
Student wealth index	0.244***	0.127*
	(0.0553)	(0.0667)
Student favorite subject Language	-0.0892	-0.183**
	(0.0702)	(0.0713)
Teacher age	-0.00273	-0.00254
	(0.00449)	(0.00480)
Teacher female	0.0159	-0.0663
	(0.0758)	(0.0770)
Teacher years of experience	0.00173	0.00429
	(0.00605)	(0.00626)
Teacher has teaching certificate	0.0406	0.0999
	(0.130)	(0.174)
Teacher speaks Spanish at home	0.198**	0.107
	(0.0903)	(0.126)
Teacher is hired by the government	-0.0387	-0.347**
	(0.0997)	(0.135)
Constant	-0.967***	0.541
	(0.274)	(0.330)
Observations	914	892

Notes: (i) teachers' collaboration is instrumented by the number of teachers per school in grades 1-3; (ii) robust standard errors are adjusted for clustering by school; (iii)*** indicates statistical significance at 1%, ** at 5% and * at 10%.

Appendix 4.D. First stage tests for sample restricted to Fang teachers

Table 4.D.1. First stage tests for regressions on sample restricted to Fang teachers

	Language	Mathematics
Endogeneity test		
Durbin- Wu-Hausman (chi squared) test	6.68001***	8.04448***
Under-identification test		
Kleibergen-Paap rk LM statistics	5.204**	6.032**
Weak identification test		
Cragg- Donald Wald F statistic	5.260	6.154
Kleibergen-Paap Wals rk F statistics	5.679	6.692
Overidentification test		
Hansen J statistics		

Notes: (i) where ***, **, * denote significance at 1%, 5% and 10%-level. (ii) Stock-Yogo weak ID test critical values: 10% = 16.38; 15% = 8.96; 20% = 6.66; 25% = 5.53

Appendix 4.E. Results Tables for Fang Teachers Sample

Table 4.E.1 Estimation of teachers' high collaboration on mathematics and language scores by 'naïve' OLS regression

Variable	Language	Mathematics
High collaboration	-0.0397	-0.0434
	(0.0971)	(0.105)
Grade	0.359***	-0.124**
	(0.0607)	(0.0627)
Student age	0.00177	0.00790
	(0.0305)	(0.0320)
Student female	0.0443	-0.0832
	(0.0900)	(0.0927)
Student household size	0.0320**	0.0102
	(0.0136)	(0.0160)
Student wealth index missing	-0.280	-0.408*
	(0.189)	(0.220)
Student wealth index	0.216***	0.136**
	(0.0656)	(0.0669)
Student favorite subject Language	-0.145	-0.266***
	(0.0949)	(0.0995)
Teacher age	0.000591	-0.00278
	(0.00558)	(0.00626)
Teacher female	-0.0796	-0.112
	(0.102)	(0.112)
Teacher years of experience	0.000574	0.00555
	(0.00782)	(0.00901)
Teacher has teaching certificate	-0.0158	0.0680
	(0.186)	(0.162)
Teacher is hired by the government	-0.286***	-0.475***
	(0.103)	(0.115)
Constant	-1.232***	0.263
	(0.344)	(0.365)
Observations	443	430
R-squared	0.214	0.116

Notes: (i) robust standard errors are adjusted for clustering by school. (ii) *** indicates statistical significance at 1%, ** at 5% and * at 10%.

Table 4.E.2. **First-stage regression results**: relationship between high levels of collaboration and total and total number of Fang teachers in primary grades

	Language	Mathematics
Number of Fang teachers in grades 1-3	0.154***	0.176***
	(0.0554)	(0.0588)
Grade	0.281***	0.289***
	(0.0840)	(0.0864)
Student age	0.00139	-0.00195
	(0.0395)	(0.0403)
Student female	0.0795	0.100
	(0.130)	(0.132)
Student household size	-0.0244	-0.0162
	(0.0211)	(0.0218)
Student wealth index missing	0.655**	0.618*
	(0.322)	(0.333)
Student wealth index	-0.466***	-0.473***
	(0.0960)	(0.0981)
Student favorite subject Language	-0.0542	-0.00633
	(0.138)	(0.141)
Teacher age	0.00505	0.00607
	(0.00974)	(0.00990)
Teacher female	0.0755	0.0525
	(0.155)	(0.158)
Teacher years of experience	-0.0255*	-0.0285**
	(0.0134)	(0.0136)
Teacher has teaching certificate	0.0223	0.0304
	(0.293)	(0.291)
Teacher is hired by the government	-0.108	-0.123
	(0.157)	(0.158)
Constant	0.615	0.538
	(0.482)	(0.495)
Observations	443	430

Notes: (i) robust standard errors are adjusted for clustering by school. (ii) *** indicates statistical significance at 1%, ** at 5% and * at 10%.

Table 4.E.3 **Two-stage least square results**: relationship between teachers' collaboration and student score in language and mathematics assessment

	Language	Mathematics
Teacher exhibit high collaboration	0.583*	0.529
	(0.342)	(0.335)
Grade	0.298***	-0.180**
	(0.0686)	(0.0711)
Student age	-0.000428	0.00579
	(0.0277)	(0.0290)
Student female	0.0295	-0.102
	(0.0924)	(0.0962)
Student household size	0.0371**	0.0136
	(0.0152)	(0.0160)
Student wealth index missing	-0.417*	-0.523**
	(0.215)	(0.226)
Student wealth index	0.304***	0.218***
	(0.0809)	(0.0831)
Student favorite subject Language	-0.135	-0.264***
	(0.0979)	(0.102)
Teacher age	9.77e-05	-0.00321
	(0.00644)	(0.00664)
Teacher female	-0.0749	-0.102
	(0.106)	(0.111)
Teacher years of experience	0.00519	0.0102
	(0.00929)	(0.00963)
Teacher has teaching certificate	-0.00236	0.0729
	(0.208)	(0.214)
Teacher is hired by the government	-0.266**	-0.451***
	(0.113)	(0.117)
Constant	-1.729***	-0.190
	(0.429)	(0.438)
Observations	443	430

Notes: (i) teachers' collaboration is instrumented by the number of Fang teachers per school in grades 1-3; (ii) robust standard errors are adjusted for clustering by school; (iii)*** indicates statistical significance at 1%, ** at 5% and * at 10%.

Appendix 4.F. Results Tables for Rural and Urban Samples

Table 4.F.1. **First-stage regression results for <u>urban sample</u>**: relationship between high levels of collaboration and total and total number of teachers in primary grades

	Language	Mathematics
Number of teachers in Grades 1 - 3	0.0275**	0.0176
	(0.0136)	(0.0115)
Grade	0.252***	0.251***
	(0.0850)	(0.0821)
Student age	0.119***	0.0902**
	(0.0396)	(0.0382)
Student female	-0.0843	-0.0782
	(0.125)	(0.120)
Student household size	-0.00611	-0.00533
	(0.0207)	(0.0209)
Student wealth index missing	0.589*	0.586*
	(0.317)	(0.345)
Student wealth index	-0.264**	-0.179
	(0.115)	(0.110)
Student favorite subject Language	0.120	0.189
	(0.133)	(0.130)
Teacher age	-0.0154	-0.0124
	(0.0101)	(0.00981)
Teacher female	-0.641***	-0.580***
	(0.167)	(0.158)
Teacher years of experience	0.0708***	0.0607***
	(0.0161)	(0.0147)
Teacher has teaching certificate	0.708***	0.651***
	(0.182)	(0.176)
Teacher speaks Spanish at home	-0.635***	-0.639***
	(0.132)	(0.126)
Teacher is hired by the government	-0.680***	-0.717***
	(0.188)	(0.187)
Constant	-0.644	-0.492
	(0.508)	(0.496)
Observations	512	495

Notes: (i) robust standard errors are adjusted for clustering by school. (ii) *** indicates statistical significance at 1%, ** at 5% and * at 10%.

Table 4.F.2. **Two-stage least square results for urban sample**: relationship between teachers' collaboration and student score in language and mathematics assessment

	Language	Mathematics
Teacher exhibit high collaboration	0.227	0.757*
	(0.377)	(0.286)
Grade	0.357***	-0.0371
	(0.0650)	(0.0758)
Student age	-0.0832***	0.00193
	(0.0288)	(0.0344)
Student female	0.133	-0.0935
	(0.0819)	(0.105)
Student household size	-0.00729	-0.0143
	(0.0138)	(0.0181)
Student wealth index missing	-0.118	0.459
	(0.223)	(0.292)
Student wealth index	0.0801	-0.0645
	(0.0798)	(0.0968)
Student favorite subject Language	-0.0768	-0.103
	(0.0899)	(0.115)
Teacher age	0.0115*	-0.00645
	(0.00684)	(0.00860)
Teacher female	0.205	-0.242
	(0.134)	(0.148)
Teacher years of experience	-0.00799	0.0379***
	(0.0114)	(0.0126)
Teacher has teaching certificate	0.184	0.570***
	(0.151)	(0.164)
Teacher speaks Spanish at home	-0.0581	-0.473***
	(0.120)	(0.126)
Teacher is hired by the government	-0.128	-0.597***
	(0.142)	(0.168)
Constant	-0.440	1.637***
	(0.359)	(0.442)
Observations	512	495
Teacher exhibit high collaboration		

Notes: (i) teachers' collaboration is instrumented by the number of Fang teachers per school in grades 1-3; (ii) robust standard errors are adjusted for clustering by school; (iii)*** indicates statistical significance at 1%, ** at 5% and * at 10%.

Table 4.F.3. **First-stage regression results for rural sample**: relationship between high levels of collaboration and total and total number of teachers in primary grades

	Language	Math
Number of teachers in Grades 1 - 3	0.00362	0.0135
	(0.0385)	(0.0363)
Grade	-0.403***	-0.424***
	(0.101)	(0.101)
Student age	0.166***	0.175***
	(0.0485)	(0.0489)
Student female	0.0526	0.0356
	(0.145)	(0.147)
Student household size	0.0320	0.0295
	(0.0264)	(0.0262)
Student wealth index missing	0.00136	0.0638
	(0.315)	(0.323)
Student wealth index	-0.441***	-0.459***
	(0.109)	(0.109)
Student favorite subject Language	-0.112	-0.0915
	(0.160)	(0.160)
Teacher age	0.0244**	0.0245***
	(0.00959)	(0.00950)
Teacher female	0.766***	0.808***
	(0.182)	(0.185)
Teacher years of experience	-0.0597***	-0.0637***
	(0.0143)	(0.0143)
Teacher has teaching certificate	0.0511	0.119
	(0.373)	(0.373)
Teacher speaks Spanish at home	-0.601***	-0.580***
	(0.166)	(0.165)
Teacher is hired by the government	-0.268	-0.243
	(0.169)	(0.177)
Constant	-0.471	-0.524
	(0.537)	(0.529)
Observations	398	393

Notes: (i) robust standard errors are adjusted for clustering by school. (ii) *** indicates statistical significance at 1%, ** at 5% and * at 10%.

Table 4.F.4. **Two-stage least square results for rural sample**: relationship between teachers' collaboration and student score in language and mathematics assessment

	Language	Mathematics
Teacher exhibit high collaboration	0.268	1.056**
	(0.622)	(0.473)
Grade	0.270***	-0.0423
	(0.100)	(0.0925)
Student age	-0.0223	-0.0148
	(0.0436)	(0.0416)
Student female	0.150	-0.0377
	(0.0952)	(0.103)
Student household size	0.0383**	-0.00358
	(0.0183)	(0.0191)
Student wealth index missing	-0.397*	-0.226
	(0.206)	(0.223)
Student wealth index	0.261**	0.323***
	(0.109)	(0.0968)
Student favorite subject Language	-0.0624	-0.102
	(0.106)	(0.113)
Teacher age	-0.00980	-0.0125*
	(0.00663)	(0.00666)
Teacher female	-0.106	-0.403***
	(0.170)	(0.151)
Teacher years of experience	0.0163	0.0240**
	(0.0122)	(0.0113)
Teacher has teaching certificate	-0.237	0.00707
	(0.243)	(0.264)
Teacher speaks Spanish at home	0.353**	0.488***
	(0.148)	(0.137)
Teacher is hired by the government	-0.201	-0.394***
	(0.125)	(0.129)
Constant	-1.058**	-0.353
	(0.425)	(0.414)
Observations	398	393

Notes: (i) teachers' collaboration is instrumented by the number of Fang teachers per school in grades 1-3; (ii) robust standard errors are adjusted for clustering by school; (iii)*** indicates statistical significance at 1%, ** at 5% and * at 10%.

Chapter 5. Do Principals' Instructional Support Practices Result in Student Achievement?[23]

5.1. Introduction

An emerging literature suggests that school principals are critical for ensuring academic achievement and in developing high-quality schools (Horng et al., 2010; Andrews and Soder, 1987). Principals can contribute to student achievement indirectly, through actions they take to influence the school and classroom conditions (Hallinger & Heck, 1999; Hallinger, 2005). Branch et al. (2012) emphasize the importance of principals, stating that "*it is widely believed that a good principal is the key to a successful school*" (p.63).

However, although there is a broad recognition that principals matter for student achievement, little is known about how and through which channels principals affect these outcomes. In contrast to the large literature on teacher quality (Araujo et al., 2016; Chetty et al., 2014; Rivkin, et al., 2005; Rockoff, 2004; Harris & Sass 2006; Kane et al., 2008; Buddin and Zamarro 2009), few studies have addressed the impact of principals in the production of student achievement (Clark et al., 2009; Horng et al., 2010; Dhuey and Smith 2014; Branch et al., 2012; Beteille et al., 2012; Coelli and Green 2012). Even fewer studies have attempted to empirically identify specific skills that principals need to promote student achievement. Exceptions include studies by Grissom and Loeb (2011), Ballou and Podgursky (1993) and Eberts and Stone (1988).

Most of the academic literature on the impact of principals on student achievement emerged in the context of developed countries, and it is unclear to what extend the findings are also applicable to developing countries. Given the uniqueness of the contexts, social norms and structural conditions found in developing countries, further research is warranted. As highlighted by Hallinger (1999), it is imperative to understand how different cultures reshape popularized concepts and practices of principals.

This paper contributes to the incipient body of empirical evidence on the influence of principals on student achievement in developing countries in general, and Sub-Saharan Africa in particular, using Equatorial Guinea as a case study. Specifically, this paper investigates to *what extend principal instructional support practices contribute to student achievement*? We provide an additional contribution to the literature by analyzing the influence of principals on mathematics and language achievement separately. Most of the available literature examined the effect of principals on composite measures of student performance, and

[23] This chapter is based on: Soares, F. & De Witte, K. Do Principals' Instructional Support Practices Result in Student Achievement?.

only a few studies have tested the effect of principals by subject area (Dhuey & Smith, 2014; Clark et al., 2009).

From an empirical perspective, determining the effect of principal instructional practices on student achievement is challenging because of omitted variable bias and the resulting endogeneity issues. Because principals who provide instructional support to teachers may be different in unobservable ways from those who do not provide instructional support or provide it less frequently, the relationship between principal instructional support and student achievement detected through a traditional Ordinary Least Square (OLS) estimation may be capturing unobservable influences on student learning that are omitted as predictors. This would violate the OLS assumption that the model error term is unrelated to the regressors (Cameron & Trivedi, 2010) and endogeneity would arise. Also, if there is non-random matching of students and principals, meaning that principals with stronger credentials work in schools with more qualified teachers and more advantaged and higher performing students, the OLS regressor becomes a biased estimate of the causal effect. There is no indication, however, that this is the case in Equatorial Guinea, which makes it a unique setting to study the effect of principals. In Equatorial Guinea, the education system is highly centralized and the Ministry of Education and Science (MoES) assigns principals to public and private schools arbitrarily without undergoing a consultation process. This process mimics a natural experiment, as principal allocation to schools and hence to students is determined arbitrarily by an outside entity, resembling a random assignment. We test empirically the hypothesis that more qualified principals work in higher performing and wealthier schools and we find evidence of randomness in the assignment of principals to schools. We argue that the random and exogenous assignment of principals to students should eliminate any omitted variable bias and endogeneity issues and the OLS regressor should be an unbiased estimate.

Another empirical challenge refers to the difficulty in separating the contributions of principal's practices from the many other factors that drive student achievement (Branch et al., 2012). Although we cannot measure unobservable factors, such as students and families' motivation, our data contains detailed characteristics on students' background characteristics, including gender, age, ethnicity, wealth level and household size. Since we lack longitudinal data to control for student's prior achievement, we use a Coarsened Exact Matching (CEM) approach to construct a pseudo-panel. Specifically, we match third grade students with first grade students based on key observable characteristics. With the CEM approach we first temporarily coarsen each key observable characteristic into unique stratas and we then exact match first and third grade students on these coarsened data, retaining only the original (un-coarsened) values of the matched data. First and third grade students with the same values *for all the coarsened variables* are placed in the same strata. Next, we generate a mean first grade score per strata and input it as the first grade score

for the third graders in that specific strata. By controlling for third grade student's estimated math and language scores in first grade, we aim to capture prior achievement of students.

There are a few reasons why Equatorial Guinea is an especially interesting setting to study the impact of principal instructional practices on student achievement. First, as discussed in more detail in chapter one, the country's education system is characterized by low teachers' qualifications, traditional pedagogical approaches and low student learning outcomes (Bassett et al., 2017). Exploring alternative and sustainable methods that can lead to improved teaching and learning is key for the country to improve the delivery of quality education. Second, in Equatorial Guinea, nearly all the principals are hired by the Ministry of Education and Science and placed arbitrarily in schools across the country, mimicking a natural experiment. Third, for initial grades of elementary schools, on which we focus, we have reliable and detailed data on student achievement in math and language as well as on student's, teacher's and principal's background characteristics. Specially in developing countries, "data suitable for doing rigorous empirical work in this area are scarce" (Grissom & Loeb, 2011). Fourth, we are able to conduct upper bound estimations. Given the poor quality of education in the country, we expect the effect of principal instruction support to be higher compared to other settings. To the best of our knowledge, this is the first paper to analyze the effect of principal's instructional practices on student achievement in Sub-Saharan Africa.

The remainder of the paper is structured as follows. In section 2, we present a summary of the literature on effect of principals on student achievement. In section 3 the Equatorial Guinea setting is briefly described, while section 4 describes the data. Section 5 addresses potential non-random matching of students and principals. Section 6 discusses the estimation strategy and Section 7 contains the empirical results. The final Section concludes with a discussion of main findings and policy implications.

5.2. Literature review

Only a small body of research empirically links principals to student achievement. While we acknowledge the limitations of focusing on student achievement in chapter three (*i.e.,* not capturing a breadth of student outcomes that individuals and society value for quality education), we consider that it may provide a highly reliable measure of student learning in the tested content. We categorize the existing economics literature related to the effect of principals on student achievement into four different strands. We add a fourth strand of the literature to the earlier proposal of Dhuey and Smith (2014). Whereas the authors focus on the overall effect of the principal on student achievement, the influence of principal observable characteristics and the impact of principal training programs, we add the literature that discusses the effect of principal leadership on student achievement.

The first strand estimates the overall effect of the principal on achievement (Branch et al., 2012; Dhuey & Smith 2014; Coelli & Green 2012). For example, Branch et al. (2012) and Dhuey and Smith (2014) use value added modeling to measure how average gains in achievement, adjusted for individual student and school characteristics, differ across principals. Both studies find that highly effective principals have the potential to raise student achievement, although the magnitude of these impacts are smaller than those associated with having a highly effective teacher (Branch et al., 2012). Dhuey and Smith (2014) analyze the effect of principals on math and reading scores separately, and although the authors find that principals have an effect on both outcomes, the magnitude of impact of an increase in principal quality is higher for mathematics (0.408 SD) than for reading (0.289 SD).

The second strand analyzes how principal observable characteristics, such as education level, years of experience, experience teaching, among others, affect achievement (Clark et al., 2009). Overall, this literature strand has found mixed results. Clark and colleagues (2009) find no relationship between principal education and school performance, measured by students' standardized exam scores and other outcomes. On the other hand, Eberts and Stone (1988) and Ballou and Podgursky (1993) find a negative correlation between school performance and principal education (more educated principals associated with worse school performance). Regarding experience, both Clark and colleagues (2009) analyze outcomes separately by subject area and find a positive relationship between principal experience and school performance for math test scores, but not for English. Eberts and Stone (1988) find a positive relationship between principal experience and school performance, but Ballou and Podgursky (1993) find no effect.

The third strand of the literature, which examines the impact of principal training programs, is relatively recent and has found some initial promising results. Using a school-level randomized field experiment, Fryer (2017) assesses the impact of a principal management training program on student achievement. The author finds that principals trained under the program significantly increase student achievement in all subjects in year one although this effect fades away in year two (which the authors attribute to principal turnover). Corcoran et al. (2012) evaluate the NYC Aspiring Principals Program and show that principals trained by APP improve school performance on standardized tests after about three years.

The fourth strand focuses on the impact of principal leadership on student achievement. Different meta-analyses of the overall influence of educational leadership (not necessarily leadership of the principal only) have shown varying results on composite measures of student achievement and have not provided more detailed evidence on specific subject areas. Witziers et al. (2003) conduct a meta-analysis using 37 multinational studies and find a very small effect size (0.02) of principals on student outcomes. Although the authors conduct the analysis on composite measures of student performance, they test if studies using math or language scores report higher/lower effect sizes than studies using composite measures, but do not

find any statistically significant difference. In a meta-analysis of 70 published and unpublished studies Marzano et al. (2005) also find a positive effect between leadership and student academic outcomes, although the effect size is much larger (0.40) than found by Witziers et al. (2003). Robinson et al. (2008) examine the impact of particular types of leadership on student outcomes. In their meta-analysis of 27 studies, they find a positive impact of transformational leadership (0.11), instructional leadership (0.42), and other types of leadership (0.30) on student outcomes. The authors conclude that instructional leadership actions are among the strongest predictors of student achievement, and that "the closer educational leaders get to the core business of teaching and learning, the more likely they are to have a positive impact on students' outcomes" (p. 664). Only one article included in the review of Robinson et al. (2008) reported results separately for math and reading. The study by Alig-Mielcarek and Hoy (2005) finds that an increase in instructional leadership is associated with an increase of 0.20 SD in math and 0.16 SD in reading.

Instructional leadership

Within the broader category of principal leadership (highlighted under strand four), the most studied aspect has been instructional leadership. Instructional leadership is a concept developed during the 1970s and 1980s as part of the effective schools' movement (Hallinger, 2005; Marks & Printy, 2003; Neumerski, 2012). This movement studied key characteristics of effective schools and claimed that effective schools almost always have strong leaders focused on instruction (Neumerski, 2012). The original concept of instructional leadership implied a focus on classroom practices and assumed that instruction would improve if principals provide detailed feedback to teachers (Louis et al., 2010). Some authors have criticized the concept as overly broad, vague, and ambiguous (Murphy, 1988). This broad view of instructional leaderships encompasses "everything a principal does during the day to support the achievement of students and the ability of teachers to teach" (Marks & Printy, 2003, p. 373).

Conceptual definitions and models of instructional leadership have evolved over time. Newer models of instructional leadership go beyond the responsibilities of principals to observe and intervene in the classroom to also incorporate responsibilities touching on vision, organizational culture, positive school climate and the like (Louis et al., 2010). The view of the principal as actively involved in curriculum and instructional issues has evolved to also incorporate the role of principals in establishing a structure and providing resources for a focus on instruction and learning in the school (OECD, 2016). This shift towards a more inclusive focus views leadership as a shared and collective effort in which not only the principal, but different school staff members, are involved (OECD, 2016; Robinson, 2008). One model proposed by Hallinger and Murphy (1985) has been widely applied and encompasses three dimensions of instructional leadership: defining the school's mission, managing the instructional program, and promoting a positive school learning climate. Our research focuses specifically on one facet of instructional leadership, related

to management of the instructional program and focused on principal's practices for instructional support and supervision. This element is particularly relevant for elementary schools, where principals tend to get more actively involved in instructional supervision (Hallinger, 2005). It also closely aligns with principals' management practices, as "instructional leaders both lead through building a mission and manage through activities that increase alignment of activities with those purposes" (Hallinger, 2005, p.229).

5.3. Context

Equatorial Guinea is a small country in West-Africa, made up of a mainland territory called Rio Muni, and five islands including Bioko, where the capital Malabo is located. The country was a Spanish colony until 1968 and had one of the highest literacy rates in all of Africa (O'Connell, 2014). After gaining independence in 1968, a 11-year authoritarian regime took place which nearly destroyed the country's infrastructure and closed its schools in 1975 (Lipski, 2004). Because the education sector was considered an opponent of the ruling regime, the country experienced the dismantling of the physical and technical-organizational infrastructure of the education sector (PRODEGE, 2011). During that time, most teachers either lost their lives and / or took the path of exile (PRODEGE, 2011). The schools were reopened shortly after a coup in 1979 that lead to the formation of a new government, but the effects of the 11-year regime are still visible (Krause, 2010). The new government also established a dictatorial rule which exists until today, responsible for the ranking of the country as one of the nine least free countries in the world (Genocide Watch, 2012). Since oil-discovery in the mid-1990s the country has become one of sub-Sahara's biggest oil producers, but a large proportion of its population still lives in poverty.

Equatorial Guinea' current education system is characterized by poor educational outcomes and inefficiencies. Gross and net enrollment ratios in primary education are low (61.58% and 43.34% respectively), indicating that a sizeable proportion of the primary school-age population is still out of school. For those who are in school, survival to the last grade of primary is low - only 72% of students who enroll in first grade are expected to reach the last grade of primary education - and repetition rates are high at 18.52%. Students in school are also not showing signs of learning – a large number of learners in the early primary grades still lack foundational reading and mathematics skills (Bassett et al., 2017).

Several factors contribute to the poor quality of education in Equatorial Guinea, including low teachers' qualification, lack of learning materials and poor physical conditions. While teacher training is improving, the rates of qualified teachers remain low. Additionally, teachers must cope with large groups of pupils of different ages all grouped together under a single class (UNESCO, 2004). Traditional classroom instructional practices tend to dominate, as teachers depend heavily on dictation and copying and make limited use of problem solving and dialogues initiated by the students (Bassett et al., 2017). The situation

is even more acute in rural areas. A report from UNESCO (2004) highlights that schools in rural areas have virtually no resources; some do not even have furniture. In rural areas, often there is no electric power, no drinking water, no latrines. There are schools that do not have their own premises, and in some villages, pupils attend class in the chapels of one of the churches.

Given the deep education challenges and inefficiencies in Equatorial Guinea, we would expect a higher effect of principal instructional support to teachers on student learning when compared to results from developed countries. This is in line with existing research which shows that effects of successful principal leadership "are considerably greater in schools that are in more difficult circumstances" (Leithwood et al., 2004, p. 5). The estimation of upper bounds, i.e. the effect of principal instruction support being higher compared to other settings, make Equatorial Guinea an interesting case study. On the other hand, we also acknowledge that there is risk that principal's instructional support may reinforce and conserve existing traditional instructional practice rather than change it.

Decision making is highly concentrated in the Ministry of Education (MEC) offices in Malabo, on the island of Bioko. Because high level MEC central office positions are used for patronage, at the policy setting level (Directors General and above) the MEC central office has high turnover and low expertise. Overall, the MEC does not have effective norms, structure, or resources in place. The MEC central office is protocol driven because it operates using norms that are mostly tradition and process oriented, submitting most decision making to formal commissions and groups or waiting for approval from the Minister. There are no mechanisms in place to appoint, hire or promote principals and teachers based on their qualifications or on the job performance (Bassett et al., 2017). The MEC central office assigns principals and teachers to public schools in an arbitrary manner without a previous consultation process with the principals or teachers that will be dispatched. There is also no indication that there is a systematic process that guide the choice of which teacher or principal is assigned to each school. Principals have no mechanisms to hire, evaluate, and promote their teaching staff based on performance, which leaves a quality 'vacuum' and the lack of autonomy and accountability regarding teacher management (Bassett et al., 2017).

5.4. Data

We use data from the Equatorial Guinea Primary Education Survey (EG-PES), conducted by the Program for National Educational Development (PRODEGE) of Equatorial Guinea in 2015. PRODEGE is a program implemented by the Non-Government Organization Family Health International 360 (FHI360) in partnership with the Ministry of Education and Science (MEC) of Equatorial Guinea. EG-PES data collection instruments were created collaboratively by members of the Education System and assessment experts. Based on pilot-testing, instruments were subsequently refined and then administered toward the

end of the school year in 2015. Instruments included student, teacher, and principal questionnaires and student assessments in language and mathematics.

The target population for EG-PES was Grades 1, 3 and 6 students enrolled in mainstream government schools. EG-PES selected these grades because they represent the beginning, middle and end of the primary education cycle in the country. A national sampling frame was constructed using data from the Education Management Information System (EMIS) of the Ministry of Education and Science. The sampling frame contained 674 primary education schools spread across the country's seven provinces. The EG-PES adopted a two-stage stratified random sampling by cluster strategy to select students and teachers from the target population. The sample is probabilistic (randomly selected), stratified by cluster, and selected in stages. The primary sampling unit (PSU) is the school (cluster) and the secondary sampling unit is the student or teacher. To ensure that schools are proportionately represented in the sample, the frame is stratified by the following variables: grade cycle offered; province; status of province capital for the two largest provinces.

EG-PES final sample consists of 66 schools: 33 schools that offer first and third grades (first cycle of primary education) and 33 schools that offer first, third and sixth grades (first and second cycle of primary education). Out of the 66 schools in the sample, 3 were no longer in operation. The other 63 schools accepted to participate in the survey during the fieldwork. 57 principals were present at the time of the data collection responded to the principal interview questionnaire. Within each school, the EG-PES collects data on 8 students per grade, totaling 480 grade 1 students, 480 grade 3 students and 240 grade 6 students, selected randomly[24]. Given the length of the assessment, in some schools the data collection team was not able to collect data on all 8 students for each grade. No student refusal was recorded.

The discussion provided under chapters two and three of this thesis show that the EG-PES complied with good standards highlighted in the survey methodology literature and identified and attempted to minimize potential survey errors where possible. As discussed in chapter 2, because the EG-PES followed best practices in sampling design it is fair to assume that the final sample represents the underlying population and that the empirical analysis conducted under the present chapter can be generalized to the population of grades 1, 3 and 6 students in Equatorial Guinea. A discussion on the development and psychometric properties of the EG-PES instruments in chapter three allowed us to provide evidence for the reliability and validity of the instruments and measures used.

The dependent variable in this study is student achievement at end of third grade. *Student achievement* is defined by two variables: the total student score in the mathematics assessment and the total score in the

[24] A Kish selection table was developed as an instrument to assist field staff in randomly selecting classrooms and students to be part of the assessment.

language assessment. Both variables are standardized with a mean of 0 and a standard deviation of 1. Our main independent variable is *instructionally supportive principals.* EG-PES teachers' questionnaire included questions designed to measure three strategic instructional interactions between teachers and principals in terms of: (1) frequency with which the principal meets with the teacher individually to discuss pedagogical topics; (2) frequency with which the principal review teacher's lesson plans; and (3) frequency with which the principal observes the teacher in the classroom. For each respondent teacher, the scores in each of the three items were added and divided by three to form a principal's instructional support scale, ranging between 1="never" to 7="weekly". An analysis of reliability reviews a Cronbach alpha of 0.73, just above the 0.70 cut-off recommended by the literature (Salvia, Ysseldyke, & Bolt, 2010). A principal's instructional support score at the school level was generated by calculating the average score from responses of individual teachers within the school. We than generated the variable *instructionally supportive principals,* which is a binary variable that indicates high engagement in instructional practices and takes the value of 1 if, on average, the principal provides instructional support to teachers weekly, twice per month or once a month and 0 otherwise (i.e., principal provides instructional support on average once every other month or less frequent than that).[25] The binary variable generated is the variable we use in our estimation strategy.

Table 5.1 presents descriptive statistics on key student, teacher and principal variables of interest. On average, approximately 47% of principals provide high instructional support to teachers. Figure 1 presents a kernel distribution of raw scores on the principal's instructional support scale. We consider principals to be highly instructionally supportive if their average score on the scale is 5 or above. Since the scores for the mathematics and language assessments were standardized, we observe a mean close to 0 and standard deviation of 1.

Table 5.2 tests if principals that provide high support to teachers on instructional topics have different observable characteristics than principals who do not provide high support to teachers on instructional topics. The results indicate no differences between the two groups across all characteristics tested, which facilitates disentangling the effect of principals' instructional support from the effect of other principals' observable characteristics.

[25] Robustness tests indicate that our findings are robust with respect to the construction of the variable. Focusing on the highly instructionally supportive principals increases the estimated marginal effect of the variable of interest (see Appendix 1). However, the results in Appendix 1 have to be carefully interpreted as there are only few (i.e. 57) highly instructionally supportive principles. This might blow up the estimated coefficients.

Table 5.1. Descriptive statistics on key student, teacher and principal variables

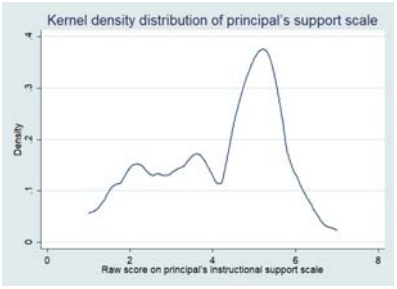
Variable	N	Mean	St. Dev.	Min	Max
Instructionally supportive principal	57	0.474	0.504	0	1
Mathematics standardized score at grade 3	294	0.003	0.986	-2.399	3.001
Language standardized score at grade 3	303	-0.006	0.979	-3.176	2.650
Female principal	57	0.421	0.498	0	1
Principal also teaches	57	0.859	0.350	0	1
Principal educational level	57	5.175	1.560	2	8
Principal experience	57	18.088	8.900	1	39
Female teacher	138	0.289	0.455	0	1
Teacher certified	138	0.123	0.330	0	1
Teacher double shift	138	0.094	0.293	0	1
Teacher years of experience	138	11.329	8.297	1	42
Teacher Fang	138	0.478	0.501	0	1
Predicted language score in 1st grade	303	0.009	0.709	-1.522	1.618
Predicted mathematics score in 1st grade	303	0.099	0.688	-1.551	-1.551
Student age	303	10.812	1.829	8	16
Student female	303	0.459	0.499	0	1
Student wealth index	303	2.010	0.681	.221	3.374
Student absenteeism	303	0.339	0.474	0	1
Student number of siblings	303	4.520	2.621	0	17
Student Fang	303	0.304	0.304	0	1

Table 5.2. Principals' background characteristics by highly supportive status

	Highly supportive principals		Non-Highly supportive principals		
	Observations	Mean	Observations	Mean	Diff
Female principal	27	0.41	30	0.43	-0.02
Principal also teaches	27	0.88	30	0.83	0.05
Principal educational level	27	5.63	30	4.77	0.86
Principal experience	27	18.88	30	17.36	1.52

*Note: Asterisks indicate statistical significance: * .10 ** .05 *** .001*

Figure 5.1. Kernel density distribution of principal's instructional support scale



5.5. Potential non-random matching issues

Potential non-random matching of students and principals may lead to biased results if principals with stronger credentials and qualifications tend to work in schools with more advantaged and higher performing students and better infra-structure. This process is called across-school sorting and has to do with how principals and students choose, or are assigned to, schools. In the case of Equatorial Guinea, the education system is highly centralized and the MEC assigns principals to public schools without undergoing a consultation process. There is also no indication that there is a systematic process that guides the choice of which principal is assign to each school. Parents in Equatorial Guinea are free to choose the school they want to enroll their children, which could introduce some bias if more motivated parents choose schools with better principals.

To test for across school sorting, we adapt the approach proposed by Clotfelter et al. (2006) and categorize principals by quality characteristics, for example, experience and education, as well as by observable demographic characteristics, such as gender. For each principal quality-demographic characteristic, the average of student and school characteristics at the school level is computed. Next, in line with Clotfelter et al. (2016) and as illustrated in Table 5.3, we run F-tests or T-tests between the means for students' score or wealth variables for the different categories of principals' qualifications and demographic profile. We also conduct an F-test or T-test to test the hypothesis that school characteristics are equal across principals' qualifications and demographic categories, as shown in Table 3. The rows of the table categorize principals in three ways (by experience, teaching degree and gender), and the columns refer to average characteristics of students at the school level or school characteristic. The table entries are means of these averages. Our goal is to test the hypothesis that student and school characteristics are equal across principals' qualifications and demographic categories. Since we do not have information on prior academic achievement for students in the third grade, we use language and math test scores data for students enrolled in first grade as a proxy for the intake of students.

Table 5.3. Evidence of Across-School Sorting: Characteristics of Students and Schools under Principals' Qualifications and Demographic Categories

Principal Qualification	Student First Grade Language Mean Score (z)	Student First Grade Math Mean Score (z)	Student Mean Wealth Index	Student Average Household Size	Total number of teachers in grades 1-3	School has Bathroom
Principal Experience						
1 to 5	-0.217	-0.200	2.091	7.14	4.641	0.64
6 to 10	0.212	0.155	2.105	6.71	6.761	0.841
11 to 15	0.129	0.197	1.868	6.28	5.195	0.75
16 or above	-0.028	-0.014	2.051	6.54	6.141	0.737
F-statistic	1.28	1.47	1.66	0.93	2.45	1.82
P-value	0.280	0.224	0.177	0.426	0.240	0.143
Principal with Teaching Degree						
No	-0.020	0.018	2.068	6.43	6.46	0.711
Yes	0.058	0.069	1.965	6.60	6.41	0.770
T-statistic	-0.685	-0.455	1.303	-0.733	0.126	-1.692
P-value	0.494	0.649	0.194	0.464	0.900	0.0912
Gender						
Female	0.163	0.106	2.157	6.34	8.06	0.708
Male	-0.075	0.005	1.88	6.68	7.14	0.791
T-statistic	-2.114	-0.922	-3.552	1.404	-0.842	-2.39
P-value	0.036	0.357	0.000	0.161	0.200	0.017

Note: For principals with a given qualification, table entries are averages of school-wide figures. Using F-tests and T-tests, the hypothesis that student characteristics are equal across principals' qualification-demographic categories is not rejected, except for the following two cases principal gender and student wealth index, and principal gender and student first grade language mean score. The hypothesis that school characteristics are equal across principals' qualification-demographic categories is not rejected, except for the following case: principal experience and school with bathroom.

Consistent with the hypothesis of non-systematic assignment of principals to schools by the Ministry of Education, table 3 shows that there is no significant differences in schools and students' characteristics across principals' qualification and demographic characteristics, with two exceptions. First, contrary to what would be expected, principals with no teaching degree tend to work in schools with higher student wealth index. In terms of magnitude, however, this difference is small. Second, principals with low experience (1 to 5 years) may be more likely to be placed in schools which do not have a bathroom, a proxy

for school's infrastructure and wealth. Overall, however, we cannot conclude that in Equatorial Guinea principals with better observable qualifications typically work in wealthier schools or schools serving higher proportions of advantaged students.

Although there is evidence of random allocation of principals across schools, principal's observable characteristics are not correlated with principal's instructional support (as illustrated in Table 2). Hence, Table 5.3 cannot test the key identification assumption of independence between a principal's instructional support and the school and students' characteristics. The absence of correlation would only be informative in testing the key independence assumption if the principals' observable characteristics were correlated with a principal's instructionally support, which is not the case in our data. In order to test the independence between a principal's instructional support and the school and students' characteristics, we run a regression of the schools and students predetermined characteristics at the school level on instructionally supportive principals. Results from Table 5.4 show no statistically significant association between teacher background characteristics, student achievement, student wealth or school characteristics with the binary variable instructionally supportive principals. Overall, our results support the independence between a principal's instructional support and key school and students' observable characteristics.

Table 5.4. Estimation of teachers, schools and student characteristics on instructionally supportive principals by Cluster Robust Ordinary Least Squares

Variable	Instructionally supportive principals
School average first grade language score	-0.012
	(0.069)
School average first grade math score	0.030
	(0.055)
School average student wealth index	-0.079
	(0.055)
Teachers certified in the school	-0.074
	(0.099)
Average years of experience of teachers in the school	0.008
	(0.006)
Total number of teachers in grades 1-3	-0.002
	(0.006)
School has bathroom	-0.028
	(0.089)
School has electricity	0.126
	(0.078)
Constant	0.119
	(0.093)
Observations	57
R-squared	0.190

Note: *** indicates statistical significance at 1%, ** at 5% and * at 10%.

5.6. Estimation Strategy

Pseudo-panel using Coarsened Exact Matching (CEM)

Given that learning is a cumulative process, student's performance observed at any given point reflects past achievements (Simone, 2013). As such, determining the impact of principal's instructional support practices on student achievement requires longitudinal data on student performance collected for the same individuals over time, which is not available for Equatorial Guinea. In order to disentangle the contribution of principal's instructional support practices to student achievement at the end of grade three, we need to control for student's prior achievement. Since our dataset contains individual level results for only one single year, we construct a pseudo-panel to predict the achievement of third graders at the end of first grade. Several previous studies have applied a pseudo-panel approach to investigate the determinants of educational achievement and human capital accumulation where longitudinal data was lacking (Brunello & Comi, 2004; Narita, 2008; Emery et al., 2012; Simone, 2013; Bleakley, 2013; Braga et al., 2013; Warunsiri & McNown, 2010; Pal & Saha, 2014).

Our research differs from previous efforts in constructing pseudo-panels as we employ a Coarsened Exact Matching (CEM) algorithm to match third grade students with first grade students based on a set of key observable characteristics. CEM is a monotonoic imbalance bounding (MIB) matching method, - which means both that the maximum imbalance between two populations may be chosen by ex-ante -, in which the data is temporarily coarsened, and then exact matches are performed on these coarsened data (Iacus, King & Porro, 2016; Blackwell et al., 2010). The idea of CEM is to temporarily coarsen each variable into substantively meaningful groups (divided into discrete categories), exact match on these coarsened data and then only retain the original (uncoarsened) values of the matched data.

Specifically, we follow the following CEM steps to construct a pseudo-panel. First, we select a set of covariates on which to match third grade students with first graders. Second, covariates are coarsened temporarily, according to CEM's automatic binning algorithm. Third, all third and first grade students are sorted into a strata on the basis of their coarsened variables. Fourth, only students with strata containing at least first grade and one third grade students are kept; others are dropped. Finally, to construct the pseudo-panel, we generate a mean first grade score per matched strata and input it as the score as first graders for the third graders in that specific strata. Since only observed variables are employed in matching, we cannot eliminate bias of omitted covariates.

We matched students based on gender, fang status, wealth index, household size, and household access to sewerage. The choice of the variables was guided by previous literature indicating these to be strong predictors of learning. Student wealth index, household size, and household access to sewerage measure

students' socio-economic background, which can strongly influence student achievement (Noel & de Broucker, 2001; Perry & McConney, 2010; PISA, 2004; Sirin, 2005). Fang status measures student's belonging to the majority and dominant Fang ethnic group, which has historically excluded and discriminated against other minority groups. Non-Fang students are more likely not to be excluded from the education system entirely and also from the learning process within schools.

The match resulted in a total of 303 third grade students (out of 399) matched to first grade students. This process generated 121 matched strata. The multivariate L1 distance prior to matching was 0.843 on un-coarsened variables and was 0.6571 after coarsening, which corresponds to a 18.59% imbalance reduction. The L1 distance provides a multivariable measure of a global imbalance. An L1 of 0 indicates perfect global balance and larger values indicate larger imbalance between groups, with a maximum of 1 (Iacus, King & Porro, 2012; Blackwell et al., 2010). The multivariate L1 distance provides only relative information (as its meaning is conditional on the data set and chosen covariates) and there are no standards or thresholds established in the literature for the match quality (Iacus, King & Porro, 2012).

Cluster Robust Ordinary Least Squares (OLS)

We estimate the effect of instructionally supportive principals on student achievement using an ordinary least squares regression model, controlling for principal, teacher and student background characteristics, including student predicted past achievement. Given the hierarchical structure of our data we employ cluster-robust standard errors that corrects the standard errors for within-cluster error correlation. It is important to control for within-cluster error correlation as failure to do so can lead to misleadingly small standard errors (Cameron and Miller, 2013). We define the cluster robust ordinary least square model as:

$$Y_{is} = \beta_0 + \beta_1 X_{is} + \beta_2 W_{is} + \cdots + \beta_{1+r} W r_i + \varepsilon_{is} \quad (1)$$

where Y_{is} is the test score for student i in school s at grade 3; X_i is the ith observation of the binary explanatory variable instructionally supportive principals; $W_{is},\ldots,W_{ri}$ are the ith observation of each of the control variables (covariates); and ε_{is} is the error term.

Random Intercept Estimate

One important assumption of the regression model (1) is that the measured units are independent, i.e., the residuals are uncorrelated with one another. If, however, the data structure is grouped, and the group effects are not taken into account in the regression model, the independence assumption will not hold (Steele, 2008). In clustered or grouped datasets, individuals from the same group might be more alike than individuals from different groups. Ignoring the clustered or grouped structure of the data may lead to

overestimated standard errors of the regression coefficients, which may in turn lead to inferring that a predictor has a 'real' effect on the outcome when in fact it does not (Steele, 2008).

Therefore, although the cluster robust OLS estimates accounts for students clustered at the school level, it does not account for students grouped in the same classroom. Given the clustered structure of our dataset, we employ a two level-random intercept model. In our data, students are nested within 70 classrooms taught by specific teachers and classrooms are nested within 54 schools, led by a principal.

5.7. Results

Table 5.5 presents the results of the cluster robust OLS estimates and random intercept estimates, examining the relationship between instructionally supportive principals and student achievement in mathematics and language at the end of grade three. Full regression tables are presented in Appendix 5.A. Both models control for achievement in first grade, obtained through a pseudo-panel approach using CEM, and other key student, teacher and principal background variables.

The results show that there is a positive and significant influence of instructionally supportive principals on the performance of students in the mathematics assessment in both specifications considered. The results show that, on average, principals that provide high support to teachers on instructional topics are associated with an increase of approximately 0.2 standard deviations in the score in the mathematics assessment. It is unlikely that this measured effect is capturing other observable characteristics of the principals that are correlated with their instructional supportiveness, since a difference in means test showed no difference between instructionally supportive and non- instructionally supportive principals on key observable characteristic (as presented earlier in Table 2). The effect found is regarded as a medium effect by the education literature. Our findings are in line with earlier literature, highlighting that teacher effects tend to be higher than principal effects in magnitude. For instance, Branch et al. (2012) found that a one standard deviation increase in principal quality translates into an increase of roughly 0.05 standard deviations in average student growth, while a one standard deviation increase in teacher quality raises achievement by somewhat more than 0.1 standard deviations. Although the magnitude tend to be smaller, it should be considered that teachers affect only their students, while principals affect all students in a school (Branch et al., 2012.)

Six covariates significantly correlate to scores in the mathematics assessment. We find a positive significant relationship between student predicted mathematics score in first grade and student mathematic achievement at the end of third grade. This is expected and consistent with the idea that learning is a cumulative process. We observe a negative significant relationship between student age and student score in the mathematics assessment. This finding is consistent with previous education research in other Sub-

Saharan African countries, which indicate that over-aged students have lower achievement when compared to regular-age students (Malpel, 2016). We observe a negative significant relationship between female student and student mathematics achievement. This is consistent with a large body of literature that have found a substantial gender gap in mathematics (suggesting lower achievement of girls when compared to boys), especially in early years of schooling (Else-Quest, Hyde, & Linn, 2010; Fryer & Levitt, 2010; Niederle & Vesterlund, 2010). We also find a negative significant relationship between student number of siblings (as a proxy for socioeconomic status) and student achievement as well as for student wealth index and student achievement. The literature indicate that socioeconomically disadvantaged students tend to perform worse on standardized measures of academic achievement compared to their wealthier peers (Noel & de Broucker, 2001; Perry & McConney, 2010; PISA, 2004; Sirin, 2005). As expected, there is also a negative significant relationship between absenteeism and student score in the mathematics assessment. Finally, we observe a negative significant relationship between teachers who teach double shift and student achievement in mathematics. Although we do not find evidence in the literature on the influence of teaching two shifts on student learning, we believe that the negative impact observed may be due to the higher workload of teachers, which may have an impact on the quality of their instruction and hence on student performance.

As observed in Table 5.5, the results do not show a significant influence of instructionally supportive principals in the language assessment. This is consistent with previous literature that analyzed student performance by subject area. For instance, Clark and colleagues (2009) find a positive relationship between principal experience and math test scores, but not between principal experience and English scores. Dhuey and Smith (2014) and Alig-Mielcarek and Hoy (2005) find a significant influence of principals in both subject areas, but the magnitude is higher for mathematics than for reading. The results are also consistent with the literature on teacher-value added. Hanushek and Rivkin (2010) review estimates of the within-school variance in teacher quality and find a lower average for reading (0.11 SD) when compared to math (0.15 SD).

Table 5.5. Estimation of instructionally supportive principals on mathematics (left) and language scores (right)

Variable	Mathematics		Language	
	OLS	Random Intercept	OLS	Random Intercept
Instructionally supportive principal	0.246*	0.225*	0.034	0.014
	(0.133)	(0.137)	(0.112)	(0.130)
Predicted score in 1st grade	0.234***	0.222***	0.174**	0.161**
	(0.081)	(0.079)	(0.080)	(0.074)
Constant	1.354**	1.171**	0.395	0.520
	(0.585)	(0.592)	(0.666)	(0.556)
Observations	293	293	303	303
R-squared	0.202		0.226	
Number of schools	54	54	54	54
Number of classrooms		70		70

Note: *** indicates statistical significance at 1%, ** at 5% and * at 10%. Full regression outcomes are provided in Appendix 5.A. The standard errors of the OLS estimates are robust and clustered at the school level. The random intercept estimates capture the differences at classroom and school level.

5.8. Heterogeneous effects by urban versus rural status

We conduct a heterogeneity analysis to identify the influence of instructionally supportive principals on student learning among schools located in urban areas versus schools located in rural areas. This analysis is especially relevant in the context of Sub-Saharan Africa, given rural-urban disparities found by previous studies in student achievement. We split our sample between schools located in urban and rural areas and estimate the same models described above for each outcome of interest (score in the mathematics assessment and score in the language assessment)

Table 5.6 presents the results for the Random Intercept estimation examining the relationship between instructionally supportive principals and student achievement in mathematics and language at the end of grade three, for both the rural and the urban samples. As expected, the results show a positive and significant influence of instructionally supportive principals in urban areas on students' score in the mathematics assessment, but no significant influence is found for principals is rural areas. Full regression tables are presented in Appendix 5.A. We control for achievement in first grade, obtained through a pseudo-panel approach using CEM, and other key student, teacher and principal background variables. Although the predicted first grade score is not significant at the 10% level under the estimations for math and language using the urban sample, it is marginally significant (p-values 0.121 and 0.111 respectively). Given poor infrastructure conditions of rural schools in Equatorial Guinea as well as the challenges teachers often face in these areas, it is not surprising that instructionally supportive principals do not have an influence on student learning. Principals designated to work in rural schools may not know how to tailor its instructional

support to address instructional challenges that might be specific to rural schools, such as over- crowded classrooms and-or diverse age groups packed into one single class.

Table 5.6. Random Intercept estimation of instructionally supportive principals on mathematics (left) and language scores (right) for urban and rural sample

	Mathematics		Language	
Variable	Urban	Rural	Urban	Rural
Instructionally supportive principal	0.348*	0.289	0.048	-0.076
	(0.179)	(0.208)	(0.165)	(0.257)
Predicted score in 1st grade	0.177	0.268**	0.117	0.193*
	(0.111)	(0.105)	(0.107)	(0.105)
Constant	1.292*	2.306***	1.379*	-0.213
	(0.739)	(0.820)	(0.712)	(0.904)
Observations	162	131	168	135
Number of schools	28	26	28	26
Number of classrooms	40	30	40	30

Note: *** indicates statistical significance at 1%, ** at 5% and * at 10%. Full regression outcomes are provided in Appendix 5.A. The random intercept estimates capture the differences at classroom and school level.

5.9. Discussion of the Results and Policy Implications

Principals have been increasingly recognized as key actors to ensuring student academic achievement and in developing high-quality schools. Nonetheless, there is a lack of rigorous empirical evidence investigating the effects of principals on student learning, particularly in the context of developing countries. Most of the academic literature on the impact of principals on student achievement emerged in developed countries, and it is unclear to what extend the findings are also applicable to developing countries. Additionally, there is no conclusive evidence in the existing literature on differential impact of principals by subject area.

This paper contributes to the literature by examining the influence of instructionally supportive principals on student achievement in developing countries in general, and Sub-Saharan Africa in particular, using Equatorial Guinea as a case study. We provide an additional contribution to the literature by analyzing the influence of principals on student mathematics and language achievement separately. We take advantage of a process that mimics a natural experiment, according to which nearly all principals in Equatorial Guinea are hired by Ministry of Education and Science and placed arbitrarily in schools across the country. We test for potential non-random matching of principals and students and we find no evidence that in Equatorial Guinea principals with better qualifications work in schools serving higher proportions of advantaged students.

Using a CEM approach to construct a pseudo-panel, third graders are matched with first grade students based on key observable characteristics. The approach allows us to generate a mean first grade score for each matched strata and hence control for third grade student's prior achievement in first grade. We than employ a cluster robust OLS estimate that accounts for students clustered at the school level, as well as a two level-random intercept estimate, which accounts for students grouped in the same classroom and in the same school.

Our initial results show that instructionally supportive principals have a positive influence on student achievement in mathematics, but not in language. Previous literature has highlighted that English test scores are less sensitive to principals' actions than math test scores, which has also been found elsewhere with respect to the impact of teachers (Clark et al., 2009). A heterogeneity analysis by urban and rural areas show a positive and significant influence of instructionally supportive principals in urban areas on students' score in the mathematics assessment, but no significant influence is found for principals is rural areas. This result is not surprising, given acute conditions of schools in rural areas, which may make it harder for principals to have an influence on student learning through instructional support to teachers.

To our knowledge, this is the first paper to construct a pseudo-panel through CEM using cross-sectional data. Our results suggest that this approach can be a suitable substitute when longitudinal data is lacking. Additionally, our results shed light into the importance of principals providing instructional support to teachers, especially in urban contexts, to influence student achievement. It also indicates that principals in rural areas may not be well-equipped to provide the type of instructional support needed to address the challenges that are specific to rural schools.

This research has important policy implications. Our findings support policies that invest in professional development activities that aim to foster principal's instructional leadership. Previous studies have found that principals who participate in instructional leadership trainings are more engaged in instructional leadership (OECD, 2016). Instructional leadership can be fostered by encouraging principals to be up to date with developments in their field through in-service training and through attendance of professional development activities. Instructional leadership trainings to principals may also have the potential to decrease the urban-rural gap in student achievement. To do so, instructional leadership trainings should be tailored to the context of rural schools and equip principals with instructional support practices that aim to tackle challenges that are specific to this context.

This research has, however, limitations that need to be highlighted. This study uses a frequency measure of instructional support and does not address the quality of this support. The distinction between task effectiveness and behavioral frequency is an important one (Grissom & Loeb, 2011). As Hallinger and

Murphy (1985) explain, ‘‘certain behaviors could be performed frequently but in a perfunctory or ritualistic manner [while] certain practices probably do not need to be performed frequently in order to be performed effectively’’ (p. 226). Without information on the quality of the instructional support provided by principals, we cannot know which specific support practices may lead to improved teaching practices and consequently increased student learning. Further research investigating the channels through which principal instructional support can lead to student learning is warranted.

References

Andrews, R. L., & Soder, R. (1987). Principal leadership and student achievement. *Educational leadership*, *44*(6), 9-11.

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

Ballou, D., & Podgursky, M. (1995). What makes a good principal? How teachers assess the performance of principals. *Economics of Education Review*, *14*(3), 243-252.

Bassett, L. K., Di Gropello, E., Marshall, J. H., Abril Tabares, J. A. (2017). *Equatorial Guinea education sector diagnostic (English)*. Washington, D.C.: World Bank Group. Available at: http://documents.worldbank.org/curated/en/356311499340208347/Equatorial-Guinea-education-sector-diagnostic

Béteille, T., Kalogrides, D., & Loeb, S. (2012). Stepping stones: Principal career paths and school outcomes. *Social Science Research*, *41*(4), 904-919.

Bleakley, H., and Hong, S. C. (2013). When the Race between Education and Technology Goes Backwards: The Postbellum Decline of White School Attendance in the Southern US. *Unpublished manuscript, University of Chicago. April.*

Blackwell, M., Iacus, S., King, G., & Porro, G. (2010). cem: Coarsened exact matching in Stata.

Braga, M., Checchi, D., and Meschi, E. (2013). Educational policies in a long-run perspective. *Economic Policy,* 28(73), 45-100.

Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2012). *Estimating the effect of leaders on public sector productivity: The case of school principals* (No. w17803). National Bureau of Economic Research.

Brunello, G., and Comi, S. (2004). Education and earnings growth: evidence from 11 European countries. *Economics of education review*, 23(1), 75-83.

Buddin, R., & Zamarro, G. (2009). Teacher qualifications and student achievement in urban elementary schools. *Journal of Urban Economics*, *66*(2), 103-115.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, *104*(9), 2593-2632.

Clark, D., Martorell, P., & Rockoff, J. (2009). School Principals and School Performance. Working Paper 38. *National Center for Analysis of longitudinal data in Education research*.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources*, *41*(4), 778–820.

Coelli, M., & Green, D. A. (2012). Leadership effects: School principals and student outcomes. *Economics of Education Review*, *31*(1), 92-109.

Corcoran, S. P., Schwartz, A. E., & Weinstein, M. (2012). Training your own: The impact of New York City's aspiring principals program on student achievement. *Educational Evaluation and Policy Analysis*, *34*(2), 232-253.

Dhuey, E., & Smith, J. (2014). How important are school principals in the production of student achievement?. *Canadian Journal of Economics/Revue canadienne d'économique*, *47*(2), 634-663.

Eberts, R. W., & Stone, J. A. (1988). Student achievement in public schools: Do principals make a difference?. *Economics of Education Review*, *7*(3), 291-299.

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: a meta-analysis. *Psychological bulletin*, *136*(1), 103.

Emery, J. (2012). Long-term consequences of natural resource booms for human capital accumulation. *ILR Review*, *65*(3), 708-734.

Fryer Jr, R. G., & Levitt, S. D. (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics*, *2*(2), 210-40.

Fryer Jr, R. G. (2017). *Management and Student Achievement: Evidence from a Randomized Field Experiment* (No. w23437). National Bureau of Economic Research.

Genocide Watch (2012). Updated country profile of Equatorial Guinea. Retrieved June 30, 2019, from http://genocidewatch.net/2013/03/19/country-profile-equatorial-guinea/

Grissom, J. A., & Loeb, S. (2011). Triangulating principal effectiveness: How perspectives of parents, teachers, and assistant principals identify the central importance of managerial skills. *American Educational Research Journal*, *48*(5), 1091-1123.

Hallinger, P. (2005). Instructional leadership and the school principal: A passing fancy that refuses to fade away. *Leadership and policy in schools*, *4*(3), 221-239.

Hallinger, P., & Heck, R. (1999). Can leadership enhance school effectiveness. *Educational management: Redefining theory, policy and practice*, 178-190.

Hallinger, P., & Murphy, J. (1985). Assessing the instructional management behavior of principals.

Elementary School Journal, 86(2), 217-247.

Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review*, *100*(2), 267-71.

Harris, D., & Sass, T. (2006). The effects of teacher training on teacher value-added. *Florida State University.* Unpublished.

Horng, E. L., Klasik, D., & Loeb, S. (2010). Principal's time use and school effectiveness. *American Journal of Education*, *116*(4), 491-523.

Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political analysis*, *20*(1), 1-24.

Leithwood, K., Seashore, K., Anderson, S., & Wahlstrom, K. (2004). Review of research: How leadership influences student learning. Available at: https://conservancy.umn.edu/bitstream/handle/11299/2035/CAREI?sequence=1

Lipski, J. (2004). The Spanish Language of Equatorial Guinea, Arizona Journal of Hispanic Cultural Studies, Vol. 8, (pp. 115-130). Tuscon, AZ: Department of Spanish and Portuguese, University of Arizona.

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education review*, *27*(6), 615-631.

Krause, J. (2010). *The Business of State Building: The Impact of Corporate Social Responsibility on State Development in Equatorial Guinea*. Dissertation, University of Florida.

Malpel, J. (2016). PASEC 2014: Education system Performance in Francophone Sub-Saharan Africa. Malpel, J. (2016). *PASEC 2014: Education system Performance in Francophone Sub-Saharan Africa*. Programme d'Analyse des Systèmes Educatifs de la CONFEMEN. Dakar, Sénégal.

Marks, H. M., & Printy, S. M. (2003). Principal leadership and school performance: An integration of transformational and instructional leadership. *Educational administration quarterly*, *39*(3), 370-397.

Marzano, R. J., Waters, T., & McNulty, B. (2005). *School leadership that works: From research to results*. Aurora, CO: ASCD and McREL

Masino, S., & Niño-Zarazúa, M. (2016). What works to improve the quality of student learning in developing countries?. *International Journal of Educational Development*, *48*, 53-65.

Murphy, J. (1988). Methodological, measurement, and conceptual problems in the study of instructional leadership. *Educational Evaluation and Policy Analysis, 10*(2), 117-139.

Narita, R. (2008). *Teenage Motherhood, Education and Labour Market Outcomes of the Mother: Evidence from Brazilian Data*. Paper presented at the Apresentação de Trabalho/Congresso.

Neumerski, C. M. (2012). Rethinking instructional leadership, a review: What do we know about principal, teacher, and coach instructional leadership, and where should we go from here?. *Educational administration quarterly*, *49*(2), 310-347.

Niederle, M., & Vesterlund, L. (2010). Explaining the gender gap in math test scores: The role of competition. *Journal of Economic Perspectives*, *24*(2), 129-44.

Noël, S., & de Broucker, P. (2001). Intergenerational inequities: A comparative analysis of the influence of parents' educational background on length of schooling and literacy skills. *In pursuit of equity in education: Using international indicators to compare equity policies*, 277-298.

O'Connel, S. S. (2014). *PRODEGE: Policy transfer to spark educational change in Equatorial Guinea*. Doctoral dissertation, University of Maryland, College Park.

OECD (2016), *School Leadership for Learning: Insights from TALIS 2013*, TALIS, OECD Publishing, Paris, https://doi.org/10.1787/9789264258341-en.

Pal, S., and Saha, B. (2014). *In Trusts We Trust: Socially Motivated Private Schools in Nepal* (IZA Discussion Papers, No. 8270). Bonn: Institute for the Study of Labor (IZA).

Perry, L. B., & McConney, A. (2010). Does the SES of the school matter? An examination of socioeconomic status and student achievement using PISA 2003. *Teachers College Record*, *112*(4), 1137-1162.

PRODEGE (Programa de Desarrollo Educativo de Guinea Ecuatorial). (2011). *Condición Docente en Guinea Ecuatorial.*

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417-458.

Robinson, V. M., Lloyd, C. A., & Rowe, K. J. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational administration quarterly*, *44*(5), 635-674.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel

data. *American Economic Review*, *94*(2), 247-252.

Salvia, J., Ysseldyke, J. E., & Bolt, S. (2010). *Assessment in special and inclusive education, 11th edition.* Belmont, CA: Wadsworth.

Simone, De G. (2013). Render unto primary the things which are primary's: Inherited and fresh learning divides in Italian lower secondary education. *Economics of Education Review*, *35*, 12-23.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of educational research*, *75*(3), 417-453.

Steele, F. (2008). Module 5: Introduction to multilevel modelling concepts. *LEMMA (Learning Environment for Multilevel Methodology and Applications), Centre for Multilevel Modelling, University of Bristol.*

UNESCO. (2004). Brunswick, E. and Valerien, J. Multigrade Schools: Improving Access In Rural Africa? UNESCO.

Warunsiri, S., and McNown, R. (2010). The returns to education in Thailand: A pseudo-panel approach. *World Development*, 38(11), 1616-1625.

Witziers, B., Bosker, R. J., & Krüger, M. L. (2003). Educational leadership and student achievement: The elusive search for an association. *Educational administration quarterly*, *39*(3), 398-425.

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

Appendix 5.A. Results Tables

Table 5.A.1. Estimation of instructionally supportive principals on mathematics and language scores by Cluster Robust Ordinary Least Squares and Random Intercept Estimates

Variable	Mathematics		Language	
	OLS	Random Intercept	OLS	Random Intercept
Instructionally supportive principal	0.246*	0.225*	0.034	0.014
	(0.133)	(0.137)	(0.112)	(0.130)
Female principal	0.110	0.111	0.039	0.020
	(0.156)	(0.143)	(0.110)	(0.126)
Principal also teaches	-0.018	-0.019	0.021	0.020
	(0.248)	(0.208)	(0.177)	(0.180)
Principal educational level	0.009	0.019	0.083**	0.088**
	(0.049)	(0.049)	(0.038)	(0.044)
Principal experience	-0.012	-0.013	-0.004	-0.004
	(0.010)	(0.009)	(0.008)	(0.008)
Female teacher	-0.023	-0.054	0.088	0.059
	(0.144)	(0.150)	(0.140)	(0.139)
Teacher certified	0.211	0.216	0.188	0.183
	(0.144)	(0.146)	(0.113)	(0.132)
Teacher double shift	0.476**	0.417*	0.262	0.255
	(0.212)	(0.216)	(0.194)	(0.201)
Teacher years of experience	-0.018	-0.015	-0.006	-0.007
	(0.013)	(0.010)	(0.009)	(0.009)
Teacher Fang	-0.010	-0.084	0.104	0.086
	(0.138)	(0.140)	(0.111)	(0.131)
Predicted score in 1st grade	0.234***	0.222***	0.174**	0.161**
	(0.081)	(0.079)	(0.080)	(0.074)
Student age	-0.120***	-0.101***	-0.132***	-0.137***
	(0.034)	(0.033)	(0.042)	(0.031)
Student female	-0.308***	-0.286***	0.003	0.012
	(0.115)	(0.104)	(0.113)	(0.104)
Student wealth index	0.035	0.022	0.255***	0.225**
	(0.100)	(0.092)	(0.084)	(0.090)
Student absenteeism	-0.181	-0.206*	0.036	-0.019
	(0.138)	(0.107)	(0.113)	(0.104)
Student number of siblings	0.039*	0.041**	0.005	0.006
	(0.020)	(0.020)	(0.019)	(0.019)
Student Fang	-0.162	-0.161	-0.278**	-0.255**
	(0.153)	(0.126)	(0.119)	(0.120)
Constant	1.354**	1.171**	0.395	0.520
	(0.585)	(0.592)	(0.666)	(0.556)
Observations	293	293	303	303
R-squared	0.202		0.226	
Number of groups		7		7

Note: *** indicates statistical significance at 1%, ** at 5% and * at 10%.

Table 5.A.2. Urban Sample - Estimation of instructionally supportive principals on mathematics and language scores by Cluster Robust Ordinary Least Squares and Random Intercept Estimates

Variable	Mathematics		Language	
	OLS	Random Intercept	OLS	Random Intercept
Instructionally supportive principal	0.336**	0.348*	0.048	0.048
	(0.163)	(0.179)	(0.157)	(0.165)
Female principal	-0.269	-0.253	-0.123	-0.123
	(0.178)	(0.189)	(0.153)	(0.173)
Principal also teaches	0.108	0.128	-0.067	-0.067
	(0.243)	(0.259)	(0.263)	(0.235)
Principal educational level	0.046	0.042	0.135*	0.135*
	(0.083)	(0.081)	(0.079)	(0.075)
Principal experience	0.005	0.005	-0.005	-0.005
	(0.009)	(0.009)	(0.007)	(0.009)
Female teacher	-0.094	-0.114	0.223	0.223
	(0.157)	(0.198)	(0.159)	(0.185)
Teacher certified	0.227	0.223	0.144	0.144
	(0.198)	(0.171)	(0.177)	(0.160)
Teacher double shift	0.612**	0.596**	0.040	0.040
	(0.296)	(0.270)	(0.363)	(0.257)
Teacher years of experience	-0.027**	-0.026**	-0.000	-0.000
	(0.011)	(0.012)	(0.012)	(0.011)
Teacher Fang	0.016	0.003	0.046	0.046
	(0.195)	(0.163)	(0.161)	(0.155)
Predicted score in 1st grade	0.177	0.177	0.117	0.117
	(0.110)	(0.111)	(0.131)	(0.107)
Student age	-0.158***	-0.149***	-0.169***	-0.169***
	(0.039)	(0.043)	(0.056)	(0.042)
Student female	-0.325*	-0.308**	0.022	0.022
	(0.165)	(0.136)	(0.148)	(0.138)
Student wealth index	0.021	0.021	-0.077	-0.077
	(0.135)	(0.138)	(0.104)	(0.138)
Student absenteeism	-0.473**	-0.471***	-0.131	-0.131
	(0.186)	(0.142)	(0.144)	(0.140)
Student number of siblings	0.027	0.027	0.014	0.014
	(0.027)	(0.026)	(0.024)	(0.025)
Student Fang	0.142	0.147	-0.080	-0.080
	(0.181)	(0.182)	(0.200)	(0.176)
Constant	1.375*	1.292*	1.379	1.379*
	(0.759)	(0.739)	(0.911)	(0.712)
Observations	162	162	168	168
R-squared	0.290		0.186	
Number of groups		28		28

Note: *** indicates statistical significance at 1%, ** at 5% and * at 10%.

Table 5.A.3. Rural Sample - Estimation of instructionally supportive principals on mathematics and language scores by Cluster Robust Ordinary Least Squares and Random Intercept Estimates

Variable	Mathematics		Language	
	OLS	Random Intercept	OLS	Random Intercept
Instructionally supportive principal	0.289	0.289	-0.040	-0.076
	(0.209)	(0.208)	(0.240)	(0.257)
Female principal	0.413*	0.413**	0.228	0.193
	(0.202)	(0.183)	(0.248)	(0.227)
Principal also teaches	-0.878***	-0.878***	0.318	0.374
	(0.265)	(0.326)	(0.306)	(0.395)
Principal educational level	-0.052	-0.052	0.077	0.079
	(0.050)	(0.057)	(0.061)	(0.071)
Principal experience	-0.023	-0.023	0.001	0.004
	(0.024)	(0.016)	(0.024)	(0.018)
Female teacher	0.403*	0.403**	0.085	0.015
	(0.217)	(0.189)	(0.249)	(0.229)
Teacher certified	0.352	0.352	0.536	0.572
	(0.497)	(0.332)	(0.456)	(0.379)
Teacher double shift	0.283	0.283	0.489	0.495
	(0.225)	(0.261)	(0.310)	(0.314)
Teacher years of experience	-0.019	-0.019	-0.003	-0.007
	(0.018)	(0.013)	(0.020)	(0.016)
Teacher Fang	-0.107	-0.107	-0.061	-0.076
	(0.365)	(0.278)	(0.414)	(0.325)
Predicted score in 1st grade	0.268**	0.268**	0.210*	0.193*
	(0.103)	(0.105)	(0.115)	(0.105)
Student age	-0.089	-0.089**	-0.128*	-0.144***
	(0.061)	(0.044)	(0.065)	(0.046)
Student female	-0.336*	-0.336**	-0.094	-0.074
	(0.192)	(0.154)	(0.191)	(0.160)
Student wealth index	-0.083	-0.083	0.405***	0.361***
	(0.131)	(0.129)	(0.109)	(0.130)
Student absenteeism	0.149	0.149	0.067	0.004
	(0.168)	(0.158)	(0.130)	(0.161)
Student number of siblings	0.049	0.049	0.002	0.004
	(0.030)	(0.031)	(0.033)	(0.031)
Student Fang	-0.538**	-0.538***	-0.287	-0.285*
	(0.221)	(0.172)	(0.182)	(0.166)
Constant	2.306**	2.306***	-0.457	-0.213
	(0.872)	(0.820)	(1.013)	(0.904)
Observations	131	131	135	135
R-squared	0.331		0.317	
Number of groups		26		26

Note: *** indicates statistical significance at 1%, ** at 5% and * at 10%.

Appendix 5.B. Robustness test of the variable of interest

In the main analysis provided in this paper, a principal is considered *instructionally supportive* if he/she provides instructional support to teachers at least once per month. Since this is a binary variable, it suggests that there is a meaningful difference between principals who fall on either side of this cut-off. Given that the design of our research dependent heavily on how we measure *instructionally supportive principals*, we construct the variable in a more conservative way and re-run the analysis using both the cluster robust OLS and the Random intercept estimates.

We construct the variable ***highly*** *instructionally supportive principals* that takes the value of 1 if, on average, the principal provides instructional support to teachers weekly or twice per month and 0 otherwise (i.e., principal provides instructional support on average once a month or less frequent than that). The difference between this and the original variable relies in the cut-off point, which is twice per month in the new variable compared to once per month in the original variable.

Results from Table A.5 show that there is a positive and significant influence of *highly instructionally supportive principals* on student performance in mathematics and language, in both specifications considered under each outcome. On average, principals that provide high support to teachers on instructional topics are associated with an increase of 0.46 standard deviations in the language assessment score and 0.59 standards deviations in the mathematics assessment score. Consistent with previous literature, the magnitude of the association found is higher for mathematics when compared to language. The difference in results on the influence of instructionally supportive principals (Table A.2) versus highly instructionally supportive principals (Table A.5) is likely due to the number of principals included under each variable. While 27 principals (out of 57 total) fall under the category of instructionally supportive, only 4 principals can be considered *highly instructionally supportive principals.*

Table 5.B.1. Estimation of highly instructionally supportive principals on mathematics and language scores by Cluster Robust Ordinary Least Squares and Random Intercept Estimates

Variable	Mathematics		Language	
	OLS	Random Intercept	OLS	Random Intercept
Highly instructionally supportive principal	0.484*	0.588*	0.405**	0.460*
	(0.247)	(0.356)	(0.182)	(0.279)
Female principal	0.078	0.081	0.018	0.000
	(0.161)	(0.147)	(0.112)	(0.126)
Principal also teaches	-0.018	-0.010	0.031	0.028
	(0.228)	(0.213)	(0.175)	(0.178)
Principal educational level	0.048	0.056	0.096**	0.100**
	(0.053)	(0.049)	(0.038)	(0.043)
Principal experience	-0.014	-0.015	-0.006	-0.005
	(0.011)	(0.009)	(0.008)	(0.008)
Female teacher	-0.044	-0.085	0.054	0.021
	(0.147)	(0.154)	(0.142)	(0.141)
Teacher certified	0.179	0.193	0.171	0.167
	(0.153)	(0.149)	(0.114)	(0.131)
Teacher double shift	0.436**	0.388*	0.290	0.297
	(0.211)	(0.217)	(0.187)	(0.197)
Teacher years of experience	-0.011	-0.008	-0.003	-0.003
	(0.013)	(0.010)	(0.009)	(0.009)
Teacher Fang	-0.067	-0.151	0.060	0.039
	(0.159)	(0.147)	(0.115)	(0.135)
Predicted language score in 1st grade	0.264***	0.239***	0.182**	0.167**
	(0.085)	(0.080)	(0.080)	(0.074)
Student age	-0.124***	-0.103***	-0.133***	-0.137***
	(0.035)	(0.033)	(0.043)	(0.031)
Student female	-0.319***	-0.290***	-0.000	0.011
	(0.114)	(0.104)	(0.113)	(0.104)
Student wealth index	-0.001	-0.004	0.231***	0.206**
	(0.102)	(0.094)	(0.085)	(0.091)
Student absenteeism	-0.202	-0.218**	0.027	-0.024
	(0.136)	(0.107)	(0.112)	(0.104)
Student number of siblings	0.041**	0.042**	0.007	0.007
	(0.020)	(0.020)	(0.019)	(0.019)
Student Fang	-0.187	-0.173	-0.292**	-0.265**
	(0.156)	(0.126)	(0.119)	(0.120)
Constant	1.450**	1.213**	0.449	0.552
	(0.597)	(0.597)	(0.677)	(0.554)
Observations	293	293	303	303
R-squared	0.196		0.230	
Number of groups		7		7

Note: *** indicates statistical significance at 1%, ** at 5% and * at 10%.

Chapter 6. Professional Learning Communities (PLC): a typology for Sub-Saharan African Countries[26]

6.1. Introduction

In efforts to improve teaching quality and promote an approach to teacher development that is both social and contextual (Du Plessis & Muzaffar, 2010), many countries have recently started to implement alternative models of teacher professional development, such as Professional Learning Communities, or PLCs. PLCs move professional development beyond the acquisition of new knowledge and skills (Vescio, Ross & Adams, 2007) to improvement of classroom practice through collaboration among teachers (DuFour, 2004). PLCs emphasize collaboration and reflection among a group of teachers, as a way to expose them to new ideas and practices so they can improve on their pedagogy through a process of critical inquiry. PLCs have become a "hot topic" in many developing countries as they hold considerable promise for teachers' capacity building for sustainable improvement in education quality (Stoll et al., 2006). Nonetheless, the literature on PLCs has typically focused on the experience of teachers in Western countries (Toole and Louis, 2002) and PLCs in developing countries have received limited attention.

PLCs in developing countries may operate in different ways reflecting their unique context, society and culture. PLCs in developed countries, for example, emphasize teachers' autonomy and authority in making decisions regarding the processes, agenda and objective of their learning communities. Previous research, however, has indicated that developing countries trying to improve very low-functioning education systems should focus on introducing highly specific approaches to instructional change (Piper et al., 2018). So, while teachers' autonomy is a key element of PLCs in developed countries, developing countries may adopt more prescriptive approaches to promoting effective instructional change. This might be needed to ensure PLCs are not reinforcing and maintaining existing traditional substandard practice rather than changing it (Stoll et al., 2006).

In a recognition that existing conceptualizations from the Western literature may not reflect how PLCs are functioning in developing countries, this research aims to inductively create a typology of PLCs that incorporates elements that might be specific to these countries, with a focus on Sub-Saharan Africa. First, we conducted a review of existing PLC conceptualizations available in the Western literature by analyzing their points of convergence, divergence, limitations and gaps. Subsequently, through a qualitative approach

[26] This chapter is based on: Soares F and Galisson K. Professional Learning Communities (PLC): a typology for Sub-Saharan African Countries. This project is supported by FHI 360.

consisting of document analysis and semi-structured interviews with experts we obtain an in-depth understanding of how PLCs are designed and implemented in Sub-Saharan Africa countries. We then integrated a typology of PLCs that emerged from our qualitative research in Sub-Saharan Africa countries with dimensions previously proposed by the Western literature into one cohesive conceptual framework. To validate and refine this framework, we held discussions with experts involved in education programs with a PLC component. The final framework can guide policy-makers and practitioners in the design and promotion of collaborative structures by accounting for elements specific to developing countries, and to Sub-Saharan African countries in particular.

6.2. Review of PLC Conceptualizations

Research on PLCs is still in early stages of theory building (Sleegers et al., 2013) and there is no broad consensus in the literature on a definition of PLC (Lomos, 2011; Vescio, Ross, & Adams, 2007; Stoll et al., 2006). However, most definitions seem to agree that PLCs involve a group of teachers sharing and critically interrogating their practice in an ongoing, reflective, collaborative, inclusive, learning-oriented, growth-promoting way (Stoll and Louis, 2007; Stoll et al., 2006; Toole and Louis, 2002; Mitchell and Sackney, 2000;). PLCs are based on the assumption that knowledge is situated in the daily experiences of teachers and best understood through critical reflection with others who share the same experience (Vescio, Ross & Adams, 2007; Buysse et al., 2003).

There is a debate in the literature on whether PLCs emerge organically in schools with effective principal and teacher leadership or if there is a place for top-down initiatives to create PLCs (Kruse & Louis, 2007). The question that remains open is: do PLCs necessarily result from spontaneous teacher action, or can education reforms or education programs proposed by local or national education authorities create effective PLCs? Earlier work of Hargreaves and Fullan (1992) has warned that collaboration is better when it is not contrived. Once PLCs are functional, however, one key point the literature highly emphasizes is teachers' empowerment and authority to make decisions regarding both the content and processes of their PLCs (Vescio et al., 2008; Huffman & Jacobson, 2003; Supovitz, 2002; Englert and Tarrant, 1995). There is also an assumption in the literature that PLCs will occur mainly within the school as the primary unit of effective change (Kruse & Louis, 2007).

There are several PLC conceptualizations available, with similarities as well as points of divergence. Most existing PLC conceptualizations incorporate three main components: (i) specific characteristics or dimensions that reflect the essence of successful PLCs (Stoll et al., 2006; Bolam et al., 2005; Bryk et al., 1999; Hord, 1997, Newmann, 1996; Kruise and Louis, 1993); (ii) factors supporting or inhibiting PLCs creation and sustenance (Atteberry & Bryk, 2011; Geijsel et al., 2009; Stoll et al., 2006; Bolam et al., 2005;

McLaughlin & Talbert, 2001; Mitchell & Sackney, 2000; Bryk et al., 1999; Kruse & Louis, 1993); and (iii) phases of development or stages of maturity of PLCs (Bolam. et al., 2005; Fullan, 1991).

Three main differences, however, underlie the existing conceptualizations. First, they consider different essential characteristics (also called dimensions) of PLCs. Appendix 6.A highlights the characteristics-dimensions considered essential by different conceptualizations established since 1993. Initial conceptualizations proposed by Kruise and Louis (1993) and Newmann (1996) focus on five key elements of PLCs: shared norms and values, focus on student learning, reflective dialogue, deprivatization of practice and collaboration. Table 6.1 provides definitions for each of these five characteristics. Later studies built upon this initial conceptualization to propose additional essential characteristics, while maintaining some or most of the original five characteristics. Although scholars may consider different essential dimensions, there is a congruence around the five essential characteristics originally proposed by Kruse and Louis (1993) and Newmann (1996), as illustrated in Appendix 6.A. There are only a few quantitative studies from Western contexts that explored empirically the five characteristics, all of which confirmed them as critical to PLCs (Bolam et al., 2005; Bryk et al., 1999; Louis & Marks, 1998).

Table 6.1. Definition of Five Essential Characteristics of Professional Learning Communities

Shared norms and values	Staff share beliefs about children and their ability to learn, beliefs about the proper roles of teachers, parents, and administrators, and beliefs concerning the use of time and space within the school.
Focus on student learning	Staff members have a sustained and undeviating focus on student learning. Focus on student learning reflects the idea that PLCs should not simply ensure that students are taught but PLCs should also ensure that students learn.
Reflective dialogue	Recurring dialogue holds practice, pedagogy and student learning under scrutiny. It implies both self-critique and institutional-critique as teachers work towards discoveries concerning their own learning and practice.
De-privatization of practice	Teachers practice and talk about teaching in public ways. It includes not only peer observation of practice and feedback, but also opened dialogue about individual teachers' practices where teachers share their successes and learn from their disappointments.
Collaboration	Involves teachers sharing expertise and working together to produce materials and activities for improved curriculum and pedagogy.

Sources: Kruise and Louis (1993) and Newmann (1996)

Second, conceptualizations differ on which supporting or inhibiting factors are considered and how they are categorized[27]. Different authors include different factors in their conceptualization and they may or may not categorize them. Categories emphasized in the literature include: external factors, internal factors, organizational factors, structural conditions, school context and school composition.

Third, the same element may be conceptualized in different ways. The same element may be considered an essential characteristic in one conceptualization and as a supportive factor in another one. For example, having mutual trust-collegial relationships is considered an essential characteristic by Stoll et al. (2006), as a supportive condition by Kruse and Louis (1993) and Hord (1997) and as an organizational factor by Bryk et al. (1999).

Limitations of existing conceptualizations

Despite the advances in PLC conceptualizations, there are a few limitations. First, the multidimensional nature of PLCs and interrelatedness of the different dimensions is not often well-conceived in the literature, contributing to limited conceptual clarity and focus (Sleegers et al., 2013). Few authors have attempted to explicitly identify how the different dimensions of PLCs are inter-related. An exception includes Huff man & Hipp (2003), who combine the re-conceptualization of Hord's (1997) model with Fullan's (1991) phases of development and map the PLC essential characteristics against the stages of development. The multidimensional nature of PLCs has also often been overlooked and only a few studies define PLCs partly in terms of educators' personal capacity, while attention has been paid to dimensions of PLCs that underlie the interpersonal capacity of educators (Sleegers et al., 2013). Exceptions include Mitchell and Sackney (2000) and Sackney et al. (2005), who address the multidimensional, multilevel nature of PLCs through focusing on three interdependent capacities - personal, interpersonal, and organizational. Sleegers et al. (2013) build on the previous work by Mitchell and Sackney (2000) and Sackney et al. (2005) to propose a conceptual model describing these three capacities and add eight underlying dimensions from the literature on PLCs. The level of personal capacity within PLCs is described by two dimensions, active and reflective construction of knowledge and currency. Dimensions underlining the interpersonal capacity include shared values and vision, collective learning, and shared practices. And finally, three dimensions underlie

[27] Factors highlighted in the literature include: national policy framework (Bolam et al., 2005), policy decisions (Stoll et al., 2006), school size, (Bryk, 1999), school leaders support PLCs and share leadership (Geijsel et al., 2009; Bolam et al., 2005; Mitchell & Sackney, 2000; Bryk et al., 1999; Kruse and Louis, 1993), trust, respect, support and inclusive membership (Atteberry & Bryk, 2011; McLaughlin & Talbert, 2001; Stoll et al., 2006; Bryk, 1999; Kruse and Luis, 1993) racial diversity, gender composition, workforce turnover (Bryk, 1999), availability of resources in the school, such as time and space to meet, information and materials (Atteberry & Bryk, 2011; Stoll et al., 2006; Mitchell & Sackney, 2000), physical proximity (Kruse & Louis, 1993), communication structures (Kruse & Louis, 1993) and individual factors, such as personal career interest, experience of stress, antipathy to change and willingness to trust colleagues (Bolam et al., 2005).

organizational capacity: 1) resources, structures and systems; 2) relationships and climate; and 3) stimulating and participative leadership.

Most of the PLC literature is conceptual in nature. There is limited validation of existing conceptualizations in high-resource contexts (exceptions may include Sleegers et al., 2013; Bolam et al., 2005; Bryk et al., 1999; Louis & Marks, 1998) and almost no validation in low-and middle-income countries. Zhang and Pang (2016) explored the characteristics of PLCs in Chinese schools as informed by the Western literature on PLCs. Lee et al. (2011) indicated that three components of Professional Learning Communities (PLCs) might be relevant in the Chinese setting: collective learning and application, supportive conditions and structures, and shared and supportive leadership.

Another limitation relies in that researchers have focused on the experience of teachers and PLCs in Western countries, most notably the United States, England and the Netherlands. The scarcity of published PLC research in English from non-English speaking countries has been acknowledged by Toole and Louis (2002). Exceptions may include a study on teacher collaboration in South Africa by Abrahams (1997); a study in Ghana on collaboration between academics and teachers, based on action research, by Pryor (1998); a paper on constructivist approaches to promote teachers learning (which included learning communities) in Mexico, by Tatto (1999); and Avalos' (1998) study in Chile of teachers participating in Teachers Professional Groups. This emerging literature, however, focuses more on the interaction between PLCs and the broader cultural, political and education context in which they are inserted, rather than on how PLCs operate (an exception being the study of Avalos', 1998). Hence, it is not possible to draw generalizations about PLC functioning in low-and middle-income countries from this literature base.

There are a number of findings from this previous research on PLCs in low-and middle-income countries that are worth highlighting. Studies by Tatto (1999) and Avalos (1998) indicate that both in Chile and Mexico PLCs did not result from spontaneous teacher action (as most commonly in developed countries), but rather emerged as part of broader educational reforms that were proposed by ministerial authorities to the schools. Despite this top-down approach, Avalos' (1998) findings suggest that the "contrived collegiality" approach can in fact lead to a culture of collaboration between the teachers. Tatto (1999), on the other hand, highlights the difficulties in developing and sustaining PLCs under Mexico's top-down structure of authority in the educational system. Similarly, Pryor (1998) mentions how the authoritarian and hierarchical systems in Ghana may influence teachers' lack of a sense of their own agency, making any critical reflection on classroom practices irrelevant. The author mentions teachers are used to being directed, and their ability to study and reflect on their own practice and also to take steps to change it, appears to be absent. Abrahams (1997) emphasize how the apartheid ideology, based on separation, segregation, inequality and authoritarianism, permeated the cultures of the schools in South Africa that took part in his

study. Some of these initial studies may indicate that "the empowering values inherent in the notion of professional learning community may conflict with a nation's most basic cultural values or recent political past" (Toole & Louis, 2002, p. 273). Another issue is that in many low-and middle-income countries, and in Sub-Saharan Africa countries in particular, the initial preparation teachers receive is highly teacher-centered, extremely didactic and focused on direct transmission of knowledge (Akyeampong et al., 2013; Altinyelken, 2010; Pryor, 1998). This may make teachers' ongoing reflection on their practices and collaboration more challenging, due to a discrepancy between the type of preparation they receive to become a teacher and the skills required for effective teachers' collaboration in a PLC.

Most existing conceptualizations consider PLCs as a school-based, organic initiative. They do not account for PLCs that are explicitly and purposefully designed by policies and programs and are part of broader educational reforms. There is limited discussion on the various ways that schools and school systems might implement a PLC in accordance with the essential characteristics. Hence, existing conceptualizations offer little guidance to practitioners and policy makers. This is problematic in the context of low-and-middle income countries, as PLCs often originate and are an integral part of education systems-policies and-or education development programs. Although the Western literature highlights a common set of essential PLC characteristics, in practice PLCs in developing countries may vary greatly in their design. This variability related to PLC designs, however, is not currently accounted for by the existing literature. Given the potential for differences in conceptualization and implementation of PLCs in developed and developing settings, and the limited literature on PLCs in less-developed contexts, this study aims to inductively create a typology of PLCs that incorporates elements that might be specific to these countries, with a focus on Sub-Saharan Africa.

6.3. Methodology

This research builds on a multi-method data collection and analysis approach, encompassing document analysis, semi-structured interviews and validation sessions. PLC document collection and content analysis formed the basis of this review. The literature acknowledges that information and insights derived from documents can make valuable contributions to a knowledge base (Bowen, 2009). However, although documents have been recognized as important sources in overall qualitative research and in theory-building specifically (Glaser & Strauss, 1967), it cannot replace other kinds of data (Atkinson & Coffey, 2004). In our case specifically, we cannot learn from documents how a PLC actually operates on a day-by-day basis. In addition, we used semi-structured interviews for triangulation – to validate our findings from the document analysis and add missing perspectives originating from the interview sources. The semi-structured interviews allowed us to explore and deepen specific aspects that emerge in the document analysis. To validate the typology we conducted discussion sessions with experts in PLCs.

Document Search

We used document analysis as a qualitative methodology for systematically reviewing and evaluating documents on PLCs in Sub-Saharan African countries. Our inclusion criteria included any project or policy document detailing the design of a PLC, with a specific focus on Sub-Saharan African countries. We included documents describing PLCs primarily at the primary education levels[28], however we did encounter examples that included both preschool and primary. We adopted a two-stage approach to search for PLC documents. Since in Sub-Saharan Africa PLCs are often part of education programs promoted by international non-profit organizations in partnership with Ministries of Education, in the first-stage we contacted key development organizations working in the education field explaining our study and soliciting documents that fell within our inclusion criteria. In a second-stage we conducted a broader internet search using specific key words[29], although documents on PLC design are less likely to be found online.

Semi-structured interviews

Upon completion of the document analysis, we conducted semi-structured interviews with four FHI 360 experts and practitioners working on the design and implementation of projects with a PLC component. The second author (KG) conducted all the interviews and took notes during the interview process. We adopted a purposive and convenience sample selection approach: (i) we focus on experts directly involved in the same programs from which many of the documents for the prior stage were drawn; and (ii) we only select FHI 360 experts, as the relationship of the authors with organization has facilitated the interview process[30]. The purpose of the interviews was to triangulate information and assess whether our findings from the document analysis was perceived similarly or differently by the experts. As part of the interview process experts reflected on their own projects/experiences and the focus was not on collecting information on a cross-cutting perspective. The semi-structured interviews also allowed us to explore in further depth specific aspects that emerged in the document analysis. Finally, through the semi-structured interviews we asked to what extent PLCs incorporate the five core characteristics highlighted in the literature. For the

[28] The choice to focus on primary education follows a shift in donor priorities in recent years from promoting school attendance to improving quality of education at early grades. PLCs have become more prevalent at the primary education level in Sub-Saharan Africa, as investments in improving teaching quality through pre- and in-service teacher professional development are often completed with a PLC component.

[29] Given there is no common agreed nomenclature for PLCs in the field, we included in our search the following key words, together with "Sub-Saharan Africa": communities of practice, teacher learning circles, communities of inquiry, professional networks, critical friends groups, study groups, teacher research collaboratives, norms of collegiality, teachers' collaboration with colleagues, professional community, learning community, and teacher networks.

[30] All experts being from FHI could potentially result in a bias. However, their participation in the semi-structured interviews was not meant to build the PLC conceptualization. Rather, their role was to give meaning to the conceptualization developed from the document review and inform us if something was missing from review. Because of the role the semi-structured interviews played in this research, it is unlikely that our sample selection approach resulted in any bias.

questions on the core characteristics, we asked the interviewees to rate the strength of the each of the five characteristics in their PLC through a series of questions using a scale of 1 – 3[31]. At the end, we were able to calculate an average score indicating the overall strength of the PLC. Appendix 6.C. provides a list of experts interviewed and the interview guide used.

Validation sessions with experts

The document analysis and semi-structured interviews informed the design of an initial PLC typology and a PLC framework. Subsequently, we conducted seven discussion sessions with nine FHI 360 experts on PLCs guided by semi-structured questions to establish theoretical validity. The idea of the sessions was to asks for experts' feedback on the accuracy of our understandings. Although the typology was informed by the experience of Sub-Saharan African countries we included experts on PLCs in Africa as well as Latin American countries in the discussion sessions to benefit from their wealth of background and practical experience on the topic. In total, we conducted six PLC validation sessions, with experts working on projects in each of the following countries: Equatorial Guinea, Ghana, Nigeria, Guatemala and El Salvador. We conducted one additional session with a senior technical staff that oversees a portfolio of projects, totaling seven meetings with nine experts. Experts participating in the sessions consented verbally before the meeting. The authors took notes of the discussions and analyzed it afterwards. Appendix 6.D. includes more detailed information on the sessions, such as the date, number of participants, and topic of discussion.

Data analysis

We adopted an inductive content analysis approach to analyze the documents and interview notes, according to which themes emerge from the data through repeated examination and comparison. Specifically, we used a two-stage analytic strategy. In the first stage we analyze the documents using an inductive approach with the goal of identifying patterns in the data by means of thematic codes (Bowen, 2009; Corbin & Strauss, 2008). We used thematic analysis in this process, recognizing patterns within the data, identifying themes that became the categories for analysis (Fereday & Muir-Cochrane, 2006). By employing an inter-comparison of documents (constant comparison), we identified and described salient patterns. In this process, we looked for substantive significance, which refers to the consistency of themes across documents (and not necessarily its frequency). As Greg and colleagues explain, "Thematic analysis moves beyond counting explicit words or phrases and focus on identifying and describing both implicit and explicit ideas within the data, that is, themes" (Greg et al., 2012, p. 10). This inductive coding approach

[31] The scale for questions related to the core characteristics as follows: The project 1) does not recognize this practice under the PLC design or promote it, 2) recognizes this practice under the PLC design but does not consistently promote it, 3) explicitly recognizes and consistently promotes this PLC practice.

allows themes to emerge directly from the data, capturing elements of PLCs that may be unique to Sub-Saharan African countries and hence not previously considered by existing conceptualizations.

The second author (KG) developed and defined codes based on close reading of the document and notes to represent the identified themes (Greg et al., 2012). The codes were discussed with the first author (FS) to check for relevance. The second author (KG) then independently reviewed each document and interview note and categorized all text using the codes. The first author (FS) reviewed the coding for a subset of randomly selected documents to check for reliability (Carey and Gelaude, 2007). Only a few discrepancies emerged, which were discussed between both authors by reviewing other documents for those themes.

In the second stage the data was abstracted further using grounded theory technique. We compared and contrasted themes, identifying structure among them (Bernard & Ryan, 1998). In this process, we analyzed how key themes were inter-related and sorted them into models (or generic categories) with similar content, generating a PLC typology. We checked the typology again against the data, to ensure that our interpretation is supported and grounded in our data (Greg et al., 2012; Bernard & Ryan, 1998).

6.4. Findings

Our document search resulted in 36 documents representing eight programs with a PLC component in seven countries in Sub-Saharan Africa: Democratic Republic of Congo, Equatorial Guinea, Ghana, Nigeria, Senegal, South Africa and Tanzania, as well as documents from the International Network for Education in Emergencies (INEE), which are available for use in many contexts. Seven out of the 36 documents were found via internet search, while the remaining were shared by the nonprofits we contacted. Four main development organizations (Family Health International 360, International Rescue Committee, Chemonics International and VVOB Education for Development), in partnership with national or local education authorities, produced the documents included in the analysis. Types of documents reviewed in the process are: PLC design documents, PLC guidelines, PLC guides or PLC manuals, project proposals, project design documents, policy briefs or white papers describing PLCs. The documents' length ranged from 1 to 282 pages with an average length of 32 pages. Most often these documents were issued by Ministries of Education or other local education authorities or by development organizations with education programs in Sub-Saharan Africa. Appendix 6.B. provides a list of the documents reviewed.

In our analysis, we found that certain PLC elements consistently appeared in most or all documents reviewed, but with a lot of variation (not in the same way). It became clear that these are important elements, but they differ based on PLCs. Table 6.2 highlights the themes that emerged from comparing shared elements across documents. The first theme is the type of grouping. It became clear from the document analysis that the PLCs' grouping (who can participate in the PLC) is a recurring theme, but PLCs are

organized in different ways in different countries: some PLCs are school-based and only include teachers from the same school while others are comprised by teachers from a cluster of geographically close schools. Within school-based or cluster-based PLCs, we also observe variation regarding teacher inclusion based on grade (some PLCs only include teachers for specific grades while others involve teachers of all grades), and subject (some PLCs are focused on mathematics or language while others do not make a distinction). Frequency of PLC meetings, meeting length (or duration) and size of PLC group were themes that also consistently appeared in most documents reviewed, but with a lot of variation (e.g., recommended meeting frequency varied from weekly to monthly; meeting duration varied from 30 minutes to 8 hours). We also found that some PLCs offer incentives to teachers' participation, although the type of incentive varies greatly (from financial incentives, such as a travel stipend, to professional incentives, such as professional development credit).

We noticed that the type of material used to guide the PLC meetings varied in their level of structure and prescriptiveness. The theme "material" emerged by analyzing the materials themselves rather than looking for recurring patterns in the text of the documents. We call scripted materials those materials that pre-define the topics of each PLC meeting, specific activities to be conducted within the meetings, and how the meetings should be facilitated. If the material is less prescriptive, and only includes step-by-step approaches or suggested topics of discussion, we classify them as "guidelines." Some of the PLCs under analysis did not have a set of materials to guide the discussion, which was left entirely to the PLC members to decide.

Facilitation was a recurring theme across the documents but appeared with variation. We found through the document analysis that some PLCs have a facilitator responsible for leading the technical discussion (who usually receives prior training), while in others, the facilitator is responsible for administrative aspects (taking attendance, registering notes, scheduling meetings, etc.). We call the former a "technical facilitator" and the later an "administrative facilitator." We also observe cases where the facilitator plays both roles, or where there is no designated facilitator and teachers in a PLC take turns in facilitation responsibilities.

Table 6.2. Main themes from PLC document analysis

Type of Grouping	• School-based (PLCs comprised by teachers from the same school) or cluster-based (PLCs comprised by teachers group of schools that are geographically nearby) • Organized by grade level, subject or mix of both
Meeting Frequency	• Weekly, bi-weekly and monthly meetings
Meeting Length	• Meetings from 0.5-8 hours
Group size	• Depends on the # of teachers in school or cluster. • Range of 2 to 70, average size is 20
Material	• Scripted materials with pre-defined topics of discussion and approach. • Meeting guidelines with step-by-step approaches or suggested topics of discussion • No set materials
Incentives	• Travel stipend (for cluster-based) • Food/refreshments • Basic supplies (notebooks, pencils, etc.) • Professional development credit
Facilitation	• Technical facilitators. • Administrative facilitators (taking attendance/notes). • Facilitators may or may not receive prior training. • No designated facilitator.
External expert	• May or may not involve specialist input in subject/pedagogical knowledge to enhance the technical discussion.
Professional development	• PLCs as stand-alone • PLCs part of a broader professional development program

Source: Own data

The semi-structured interviews confirmed the themes and patterns found in the initial document analysis. Table 6.3 summarizes the findings from the semi-structured interviews. During the interviews, we explored the main themes from the document analysis (Table 6.2 above) in each of the four project examples. This data also helped us to confirm the classification of each PLC model. In addition to showing the variation across the themes listed above, Table 6.3 shows the overall score for the core characteristics of each PLC example. All four examples were in a mid-range of 2, in which meant they recognize a particular practice linked to the core characteristic, but are not consistently implementing it in the PLC. This could also be due to the fact that all four of the PLCs examined were relatively new and part of teacher development projects that brought the PLC component to teachers' practice.

Table 6.3. PLC classification based on document analysis and semi-structured interviews for four projects

	Ghana Learning	**Nigeria RANA**	**Equatorial Guinea, PRODEGE**	**Senegal, Passerelles**
PLC terminology	School-based INSET meetings	Weekly school meetings	Teacher circles/networks	Pedagogical cluster
Core characteristics[32]				
Shared norms and values	2.6	2	3	3
Focus on student learning	2	1.5	2.5	2.5
Reflective dialogue	2.5	2	2	2.5
Deprivatization of practice	2.6	2.6	2.3	2.3
Collaboration	2	1.6	1.3	2.3
Average score (1=low, 3=high)	**2.3**	**1.9**	**2.2**	**2.5**
Structural varying features				
Type of Grouping	School-based	School-based	Cluster-based	School-based and Cluster-based
Meeting Frequency	Weekly	Weekly	Bi-monthly	Monthly
Meeting Length	1 hr	1 hr	4 hrs	1-2 hrs
Group size (average)	4-10	4-8	20-30	Varies
Material	Scripted materials Meeting guidelines[33]	Meeting guidelines Meeting log	Scripted materials	Meeting guidelines Coaching rubric
Incentives	Professional development credit	None	• Travel stipend • Basic supplies • Professional development credit	None
Facilitation	• Technical and administrative facilitator • No training • Head teacher or volunteer teacher	• Technical and administrative facilitator • Receives training • Lead teacher	• Technical and administrative facilitator • Receives training • Experienced teacher	• Technical and administrative facilitator • Receives training • Experienced teacher or director
External expert	None	None	None	For specific subject content area

[32] The scale for questions related to the core characteristics is as follows: The project 1) does not recognize this practice under the PLC design or promote it, 2) recognizes this practice under the PLC design but does not consistently promote it, 3) explicitly recognizes and consistently promotes this PLC practice.

[33] The Ghana Learning Project team began implementing the PLC component with scripted materials, but found they weren't being used consistently by teachers due to their rigid structure. In 2019, later in the project, the team revised the PLC materials so that they provided teachers with more autonomy to choose their own agenda topics.

Professional development	Part of a broader TPD program	Part of a broader TPD program	Part of a broader TPD program	Part of a broader TPD program
Supporting Conditions				
	• Leadership support	• Leadership support • Built in meeting time	• Leadership support	• Leadership support • Built in meeting time
Phases of development				
	Implementation (structure is institutionalized in system)	Implementation	Implementation	Initiation[34]
Model				
	Scripted to structured	Structured	Scripted	Structured

Source: Own data

[34] The Passerelles Project was just initiating at the time of the interview. Answers are based on the intended design of the PLC and are not a reflection of what has been occurring. Thus, the PLC is still in an "initiation" phase of development even though the core characteristics reflected in the design are strong.

Interviewees confirmed to know about variation on the type of groupings. Half of the examples they offered had school-based meetings, and half had cluster-based meetings. Teachers in all the PLC examples from the interviews were grouped together by their participation in the broader teacher development program, but the composition of the groups varied similar to what we saw in the document analysis. For example, all teachers in grades 1-3 in the Nigeria Reading and Numeracy Activity (RANA) program, who received training in literacy and numeracy through the program, participated together in the weekly meetings. In the PLCs promoted by the Program for Educational Development of Equatorial Guinea (PRODEGE), all primary and preschool teachers, regardless of grade-level, were grouped together in geographic clusters as part of a national pedagogical certificate program. In terms of the meeting frequency and length of meeting, the interviews revealed variation across the examples similar to what we found in the document analysis. From the interviews, we find that the less frequent the meetings (monthly or bi-monthly as opposed to weekly) led to longer meeting times. Equatorial Guinea's bimonthly PLC meetings are four hours long, while Ghana, Nigeria and Senegal promote one to two-hour weekly meetings. Group size also varied greatly across the examples from the interviews and were largely driven by the number of teachers participating the Teacher Professional Development (TPD) program in a school or geographic area.

The interviews allowed us to deepen our understanding of the use of "materials". During the interviews we confirmed the types of materials used and determined the level of structure and whether or not a material can be categorized as "scripted." Using the categories of "materials" from the document analysis, we find that each example from the semi-structured interviews used either a material that provided basic guidelines for the meeting, such as in Nigeria and Senegal; or a heavily scripted material, such as in the case of Ghana and Equatorial Guinea. In terms of the "facilitation" characteristic, all the PLC examples from the interviews did have a facilitator, but the type of facilitator and level of training varied as in the examples from the document analysis. For example, in Ghana, the facilitators received no prior training on their facilitation role; while in Nigeria, Senegal, and Equatorial Guinea, the facilitators all received training through the projects. The designation of facilitator also varied across all examples, from head teachers to lead teachers to a volunteer teacher from the group, which confirms what we find in the document analysis.

Similarly, the incentive varied across the examples. The interviews allowed us to understand more deeply the influence of the incentive or lack of incentive on the PLC model as a supporting condition. Knowing what supporting conditions can foster teachers' participation in PLCs is helpful for Ministries of Education, donors and education practitioners to consider when designing education reforms or teacher professional development programs. For example, in Ghana, the interviewee spoke of the importance of receiving professional development credit for initiating PLCs and keeping the teachers motivated to continue their participation. In Equatorial Guinea, the teachers receive several incentives (travel stipends, professional

development credit, and basic supplies) in order to encourage participation in the PLCs. On the other end of the spectrum, in Nigeria and Senegal, no incentives are provided, and the lack of an incentive was not mentioned as one of the barriers or challenges to the teachers' consistent participation in the PLCs. Thus, the variation of incentives found in the interviews confirmed the variation found in the document analysis. The findings seem to indicate that offering an incentive at the onset might help to establish or create a demand for PLCs in an education system. However, if the incentive is removed, the PLC might end if the teachers do not have the intrinsic motivation to meet. More research would be needed to determine the impact of incentives on the establishing, implementing and sustaining PLCs over time. It would be particularly valuable to examine this element more closely in the Nigeria and Senegal examples, in which no incentives were provided.

The semi-structured interviews also allow us to explore in in further the supporting conditions and barriers for establishing and sustaining PLCs. In the case of Senegal, the interviewee stressed the importance of building positive interpersonal relationships with the teachers as an enabling factor for establishing as well as sustaining a PLCs. The document analysis, specifically the training materials and PLC guides from various projects, also revealed that many of the programs stressed the importance of building a cohesive peer support group from the onset by incorporating specific team-building exercises in the facilitator training or PLC meetings during the first sessions. Another interviewee stressed the importance of building in enough time for training the facilitators in order to ensure the success of PLCs. The training needs to be practical and purposeful for teachers to be able to lead successful meetings, particularly under a non-scripted model, as in the case of Nigeria.

In the second stage, we abstracted the data further to identify key determinants of the internal PLC dynamic and of teachers' level of autonomy. We compared and sorted a few themes into three emergent conceptual categories of PLCs. We defined three models of PLCs that can explain practices in Sub-Saharan Africa, namely the *Autonomous* PLC, the *Structured* PLC and the *Scripted* PLC.

Table 6.4. Emerging typology: three models of PLC

Autonomous PLCs.	Material (no external materials provided). Incentives (no incentives). Professional development (stand-alone).
Structured PLCs.	Material (structured steps; general guidelines). Incentives (stipend/incentive may be provided). Professional Development (may be stand-alone or part of a broader professional development program).
Scripted PLCs.	Material (highly scripted guides). Incentives (stipend/incentive may be provided). Professional Development (part of a broader professional development program).

Source: Own data

According to the *autonomous* model teachers identify their own starting points, learning needs, the focus and objectives of their learning (so they can focus on issues important to them), as well as the pace and scope of the PLC. *Autonomous* models of PLC tend to emerge organically and naturally within schools (hence are not part of structured professional development interventions). They do not rely on external materials to guide or dictate the discussions and do not provide incentives to teachers, who are self-motivated to participate. The documents analyzed did not contain an example of a fully *autonomous* model. Nonetheless, descriptions of how PLCs function from the Western literature indicate that this model is dominant in high-resource contexts. Indeed, the literature highly emphasize the importance of teachers' empowerment and authority to make decisions regarding both the content and processes their communities for PLCs to be successful (Vescio et al., 2008; Huffman & Jacobson, 2003; Supovitz, 2002; Englert and Tarrant, 1995). As such, we include it in our typology.

In the *structured* PLC model, teachers receive materials highlighting a structure or series of steps to guide their dialogue within the learning community. Teachers may or may not receive financial or other sorts of incentive to participate. *Structured* PLCs can be a stand-alone intervention (e.g., promoted by a Ministry of Education or other local education authority through the provision of guidelines), or they may be part of a professional development initiative, where the PLCs serve to reinforce and contextualize knowledge-practices taught as part of the broader program. In this model, teachers maintain some level of autonomy, but within the proposed structure. PLCs in Nigeria, for example, follow a six-step approach to each 60-minute meeting, consisting of:

1) Opening: Letter sound movement or song (3 minutes)
2) Discussion: A weekly success (2 minutes)
3) Discussion: A weekly challenge (5 minutes)
4) Practice: One instructional skill (letter blending, reading aloud, etc.) (10 minutes)
5) Practice lesson for the upcoming week (30 minutes)
6) Feedback on practice lesson (5 minutes)

Within this proposed structure, teachers can choose the topics for weekly successes and challenges and which instructional skills they would like to practice. PLCs in South Africa also promote a structure, but for a series of meetings, as opposed to each PLC meeting (Table 6.5). PLCs start with meetings focused on access and motivation and gradually, as members start to get to know each other, move to meetings focused on constructing knowledge together and critically reflecting and inquiring.

Table 6.5. PLC structure in South Africa:

		Stage 2	Stage 3	Stage 4	Stage 5
Focus	Access & Motivation	Socialization	Information Exchange	Knowledge Construction	Development (Reflection)
Approx. duration	2 meetings	2 meetings	4 meetings	Variable	Variable
Activities	- Facilitating access - Icebreakers - Discussing rules	- Getting to know each other. - Develop shared understanding on PLCs. - Developing mission & vision	- Presenting and discussing results of individual tasks. - Exchanging teaching resources	- Discuss questions for enquiry - Development of resources in effective work teams. - Lesson study - Error analysis	- Exploring the use of new materials and technologies

Source: Professional Learning Communities: A Guideline for South African Schools

According to the third PLC model, *scripted*, PLCs receive scripted materials that pre-define the objective, focus, topics, learning exercises and dynamic of the meetings and sometimes frequency with which meetings should take place. In this model materials are stand-alone so that teachers can facilitate their own learning and reflection using the scripts provided. *Scripted* PLCs reinforce knowledge-practices taught by a broader professional development program or act as channel for promoting instruction associated to curriculum reforms. As with *structured* PLCs, they may or may not provide incentives for participation. Equatorial Guinea and Ghana are countries where *scripted* PLCs are practiced. In Equatorial Guinea, the Program for Educational Development (PRODEGE) developed scripted guides that dictate the agenda and activities for the four-hour bi-monthly meetings. The PLCs are part of a national teacher professional development program in which the preschool and primary teachers who successfully complete the requirements will receive a pedagogical aptitude certificate. Given the requirements of the Teacher Professional Development (TPD) program, all the content for the meetings is pre-defined in order to cover the required content of the program which is divided into four thematic areas. For example, the guides for the first thematic area of the TPD program guide the teachers through a series of scripted activities related to basic pedagogical theory that supports active learning methodology, which is the backbone of the PRODEGE intervention. The three thematic areas that follow are subject-specific - pedagogy of mathematics, language and social and natural sciences. Since teachers are grouped by either preschool-level or primary-level, PRODEGE developed specific guides for each education level following the same thematic area. The types of activities in the guides vary from reading a text and discussing what was learned in groups or pairs, to sharing related practical classroom examples with peers and planning together how the teachers will apply the new knowledge in their own classrooms. The hope is that the certificate will motivate teachers to participate in the PLCs, connect teachers who were previously isolated, and create peer support networks nationwide that can be sustained beyond the project.

Similarly, in Ghana, the Learning project initially provided scripted materials that dictate the topic, content and timing of each PLC meeting. The Ghana materials developed for PLC meetings define a specific topic related to early grade reading instruction for each week, and lead teachers through activities, such as discussion, planning, with their suggested timing for each hour-long meeting. The content reinforced the training in early grade literacy that teachers had received and helped facilitate discussion on the new methodology. The meeting guides offered prompts for facilitators to follow in order to help guide each meeting. Given the highly scripted nature of these materials used at the onset of the project, Ghana PLCs can be categorized as *scripted.* However, the semi-structured interview revealed that changes between models can occur in practice. During implementation of the Ghana Learning PLC model, the project team found that many teachers did not participate in the meetings or did not use the meeting materials consistently. Thus, later in the project cycle, Learning released revised PLC materials for meetings that took on a more structured approach in order to give teachers more flexibility in choosing the topics most relevant to them for discussion with peers. This change signals a move from the *scripted* model to a more *structured* model.

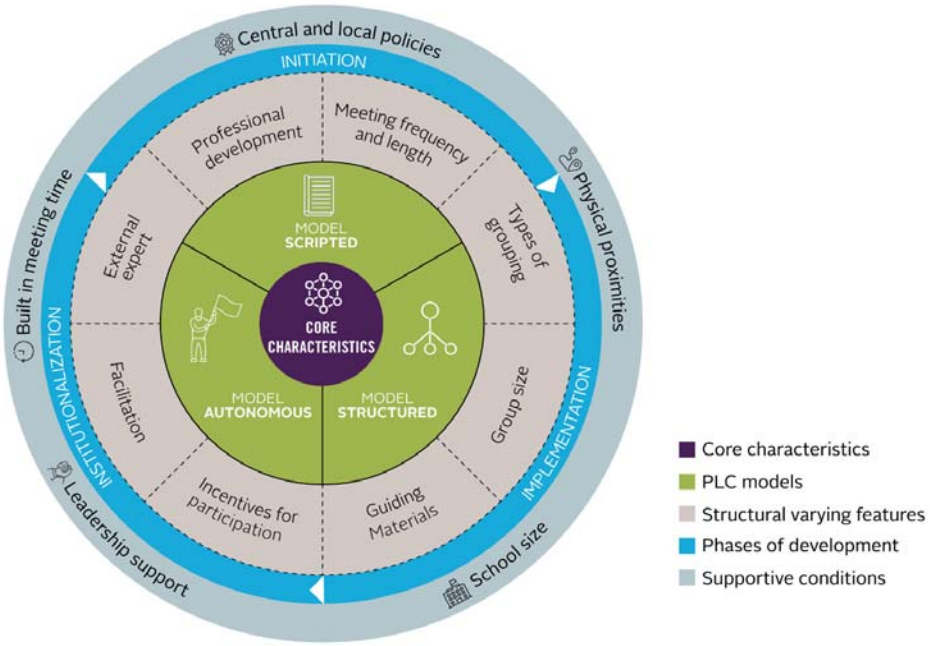
6.5. An integrated conceptual framework of PLCs

We developed a conceptual framework of PLCs that integrates the PLC typology and PLC varying features that emerged from our qualitative research in Sub-Saharan Africa countries with key elements proposed by the Western literature. We propose a five-dimension framework that integrates the three dimensions from the academic literature (core characteristics, supportive conditions and stages of maturity) with two new dimensions found in our qualitative research in Sub-Saharan Africa countries research (PLC model and structural varying features). These dimensions can be visualized in Figure 6.1 as the five circles. While each structural varying feature may or may not be important in explaining PLCs in Sub-Saharan Africa, the type of model is essential to understanding their design and operation.

At the very center of the framework are the *core characteristics of PLCs*, which we describe as the characteristics any PLC should have to be successful, regardless of the model or specific design it adopts. They refer to the five essential characteristics that received the most support in the Western PLC literature: shared norms and values; focus on student learning, reflective dialogue, de-privatization of practice, and collaboration. Although in our document analysis we do not find much evidence of the five core characteristics in Sub-Saharan Africa PLCs (neither together or separately), in the semi-structured interviews we asked to what extent PLCs incorporate the five characteristics. The interviews revealed that these characteristics are being implicitly considered in the design and implementation of the PLCs, however none of the examples promoted all five characteristics together (see Table 6.3. for these results). We

maintain the five characteristics in the framework but acknowledge that further empirical tests in the context of low-and middle-income countries are needed to better understand their role in PLCs in these countries.

Figure 6.1. Framework of Professional Learning Communities



PLC models, which emerged from our qualitative research, are directly related to the types of materials and the extent to which these materials influence the level of PLC autonomy and structure. By categorizing PLCs into three different models and acknowledging their variability in this aspect, we are able to incorporate modes of PLC operation that are specific to Sub-Saharan African countries, and possibly to low- and middle-income countries more generally. The *autonomous* model, which emphasizes teachers' autonomy and authority in making decisions regarding the processes, agenda and objective of their learning communities, is the dominant model in developed countries. The *autonomous* model, however, does not reflect how PLCs operate in Sub-Saharan African countries, where highly structured or scripted learning communities guided by pre-developed materials prevail, as we found in our document analysis.

While the three models - *scripted*, *structured* and *autonomous* - are defined by their overall level of structure and outside influence on the community (usually through materials), within any model specific features of

PLCs can vary greatly. The third dimension of our framework refer to *structural varying features*, - the different elements associated with how PLCs are designed that might vary across PLC models. Our document analysis of PLCs in Sub-Saharan African countries indeed revealed great variation. The features included in our framework refer to the themes that emerged from the document analysis, presented in Table 2: (i) *type of grouping;* (ii) *meeting frequency and length*; (iii) *group size*; (iv) *guiding materials;* (v) *facilitation*; (vi) *incentives for participation*; (vii) *professional development*; and (viii) *external expert.*

The fourth dimension of the framework refers to the *phases of development* and reflects PLCs' evolving and fluid nature, rather than static nature (Stoll et al., 2006; Fullan, 1991). It recognizes that sustainability of PLCs within schools or among clusters of schools depends on how well staff can sustain their efforts and how embedded PLCs are into the culture of their schools (or cluster of schools) (Huff man & Hipp, 2003). Specifically, we incorporate three phases of change proposed by Fullan (1991): initiation, implementation and institutionalization. This dimension is directly related to the PLC core characteristics of our framework and reflect their level of development and institutionalization within a school or cluster of schools. Some schools or clusters of schools may be at very early stages of developing the characteristics of a PLC or may be further along in the implementation process. In an institutionalized PLC, the five characteristics become embedded into the culture of the school or cluster of schools.

As part of the fifth dimension we consider different *supportive conditions* related to organizational factors, structural conditions, and overall context. We define supportive conditions as external factors that may enable or inhibit the creation, development, ongoing management and sustainability of PLCs (Bolam et al., 2005). As such, although we recognize that internal PLC characteristics may foster or hinder PLC work (e.g., mutual trust, PLC composition related diversity and background), in this dimension we focus on external factors only. We selected factors from the literature that seem particularly relevant to the Sub-Saharan African context, in light of the document analysis and semi-structured interviews. A first supportive condition is *built in meeting time*: if teachers do not have meeting time built into the school calendar, there is a risk that meetings or other forms of collaboration will be canceled or will not take place. Time for PLCs' meetings "cannot be simply tacked onto the ends of already tiring school days" (Kruse & Louis, 1993, p. 16) and teachers must be able to consistently allocate time to such meetings. Building time into the school calendar for PLCs to meet can result from school-level organization decisions or enactment of policies at the system-level. A second condition is *physical proximity and-or devoted space.* Teachers that are part of cluster-based PLCs may find it easier and may be more motivated to travel and meet if the meeting location is close by. For school-based PLCs, having physical spaces devoted for faculty meetings may increase teacher contact and minimize interruptions. A third often-cited condition is *leadership support.* Principal and headteachers play a key role in creating a school culture that fosters and generates

opportunities for collaboration among teachers and continuous learning (although they cannot ensure it will happen). A fourth condition, that may be especially relevant for low- and middle- income countries is *school size*. Small schools can be engaging work environments that facilitate the communication flow (Bryk et al., 1999). Schools that are too small, however, may not have enough teachers to collaborate in school-based PLCs. This may be the case of some multigrade schools in rural parts of Sub-Saharan Africa. The final condition refers to *central and local policies*. Central and local policies may provide incentives for Continued Professional Development (CDP) or credit towards it, may affect resource allocation to schools, may or may not allocate a certain number of hours for teachers' planning and collaboration, and may change the focus away from collaboration by placing emphasis on performance measured by student achievement.

Validation sessions with experts

After developing this draft framework, we conducted discussions with expert staff to solicit their feedback on the "fit" of the framework to the context in which they work. Overall, experts agreed with the framework developed, validating the thinking and categories proposed by the researchers. School leadership support was a condition highly emphasized by experts as a critical one for enabling and sustaining PLCs in Sub-Saharan Africa. One expert mentioned that "school principals play a critical role in supporting teachers' participation within PLCs".

Discussions around how phases of development may relate to the PLC model and some varying features, Experts highlighted that for initiating PLCs a scripted PLC model guiding participants may be beneficial, while the ultimate goal for sustainability purposes would be for PLCs to become fully autonomous in order to be institutionalized. Experts also mentioned that in terms of the type of grouping, cluster-based PLCs that build upon an established education system's organizational structure (e.g., schools already grouped together for administrative purposes) may help sustain the PLC without the outside support of a project or professional development program.

One element not previously considered in the framework was brought up. According to one expert: "we should consider adding to the typology what activity occurs during the PLC meetings. For example, do teachers use the time to look at student work, develop lesson plans, model lessons, conduct peer observation, or is time used just for reflection?" While the feature "Activities" did not emerge from the document and interview analysis, further research is needed to investigate if indeed this a relevant component of PLCs that should be integrated into the framework.

6.6. Discussion

In this paper we propose a typology of PLCs' that captures three PLC models – *autonomous, structured and scripted* – accounting for modes of PLC operation in Sub-Saharan African countries. Although we aimed to develop a typology that reflects how PLCs function in low- and middle-income countries, and in Sub-Saharan African countries in particular, the final typology also holds for developed countries. While in low- and middle-income countries the *structured* and *scripted* models of PLC prevail, in developed countries the *autonomous* model is more commonly found. Indeed, the Western literature highly emphasizes teachers' autonomy and the importance of teachers' ability to make decisions regarding the processes of their learning communities. We can also find *structured* models of PLCs in developed countries, although less frequently. For example, Hollins et al. (2004) report a structured dialogue problem-solving approach to PLCs in the United States.

Our document analysis revealed that within any of the three models specific features of PLCs can vary greatly, so we also account for *structural varying features*, which refer to all the distinctive features associated with how PLCs are designed that might vary across PLCs. There are also key differences between high and low-resource contexts in terms of *structural varying features* that the typology incorporates. For instance, while PLCs are described in the literature almost exclusively as a school-based initiative, in developing countries they often originate and are an integral part of education systems and-or education development programs.

We integrate our typology of PLCs and corresponding structural varying features with key elements proposed by the Western literature, forming one cohesive conceptual framework. The framework is organized around five interrelated PLC dimensions, three of which emerge from the Western literature (*core characteristics*, *supportive conditions* and *phases of development)* and two (PLC *models and structural varying features*) are based on a document review of PLCs in low- and middle-income countries and semi-structured interviews with experts, validated through expert discussion sessions. Our purpose with this framework is to improve conceptual clarity around the multidimensional nature of PLCs and guide policy-makers and practitioners on the design and promotion of these collaborative structures.

This study considers the experience of a sample of Sub-Saharan African countries with a focus on primary education. In order to expand the validity of our typology into other settings, more research is needed exploring other regions of the world and other educational levels (e.g., secondary and tertiary). Nonetheless, we believe this study provides further insights into how PLCs are functioning in developing countries and support to the development of PLC's conceptual foundations accounting for the reality in these countries.

Finally, by proposing specific models according to which we can classify PLCs in a clear and consistent manner, we open the doors for comparative analysis across different PLCs and different settings.

We acknowledge a few limitations of this study and final framework. First, the documents reviewed may not include the universe of existing documents on PLCs in Sub-Saharan Africa. Local, regional or even national governments in Sub-Saharan Africa might have developed documents on PLCs independently, without the support of international organizations, that are not available online. Additionally, only four development organizations shared documents on PLCs. This may be explained by the fact that PLCs are a relatively new approach in the field of international education. Second, we adopted a convenience sample selection approach to the semi-structured interviews and validation sessions. Because we only included FHI360 experts, we may be missing potentially different views and perspectives of experts working in other organizations. Finally, although we include the five core characteristics highlighted by the Western literature in our framework, there is a lack of studies investigating their relevance in low-resources contexts (an exception being Zhang and Pang, 2016) and even the empirical evidence-base in high-resources contexts is limited (exceptions include Bolam et al., 2005; Bryk et al., 1999; Louis & Marks, 1998). Further research is warranted to investigate the extent to which these characteristics are reflective of the experience and functioning of PLCs in low- and middle-income countries.

Despite these limitations, the proposed integrated framework offers opportunity for future research to disentangle the relationships between the five dimensions and test empirically how these dimensions are inter-related. We need to better understand the role played by each dimension in promoting the five core characteristics in order to design strong PLCs. For instance, which of the three PLC models is most effective in promoting the five PLC core characteristics? And does this relationship between PLC model and core characteristics vary according to phases of development - is it more beneficial to have *scripted* approaches when initiating a PLC versus an *autonomous* model for more mature, stronger PLCs? The *structural varying features* may also play an important role in strengthening the five core characteristics. For example, does having a trained facilitator can stimulate (or hinder) the development of any of the five core characteristics?

The framework also allows for more research on which PLC model can be most effective in producing different types of instructional change. The literature distinguishes between reformation versus transformation of practice (Lee et al., 2018; Philpott & Oates, 2017; Servage, 2008). Reformation of practice involves aligning individual practices with prescribed norms or existing mandated goals (Philpott & Oates, 2017). Transformation of practice involves questioning goals and empowering teachers to be "… producers of pedagogical knowledge rather than solely consumers and implementers" (Philpott & Oates, 2017, p. 213). We need to further understand if scripted PLCs can effectively lead to transformation of

practice or if they are more likely to work as vehicle for implementation of authority-endorsed change or mandates (Lee at al., 2018).

Finally, since PLCs are imbedded in a social, political, cultural and economic context, the typology opens the debate on which PLC model is best for different settings. In contexts characterized by high power distance, hierarchy and authoritarian education systems, which PLC model can more effectively facilitate changes in instructional practice? In such contexts, does the scripted model reinforces existing power dynamics or can it be helpful in moving towards real collaborative structures? Another question arises on which PLC model is more suited for different levels of teacher capacity. When there is low teacher instructional capacity and-or low teacher ability to actively and reflectively construct knowledge about teaching and student learning, which PLC model can be more helpful in filling these gaps?There are still many questions to be answered around the optimal design and implementation of PLCs in different contexts. The new empirically-based typology and integrated conceptual framework presented in this study provides a way to classify PLCs in a clear and consistent manner to facilitate future comparative analyses - across different PLCs and different contexts. By building on this foundation, we can more efficiently and effectively realize the potential of PLCs to strengthen teachers' instructional practices and improve educational outcomes.

References

Abrahams, M. (1997). Collaborative space in South African schools: A comparative perspective. Journal of Negro Education, 66(4), 409-422.

Altinyelken, H. K. (2010). Pedagogical renewal in sub- Saharan Africa: the case of Uganda. *Comparative Education*, *46*(2), 151–171.

Akyeampong, K., Lussier, K., Pryor, J., & Westbrook, J. (2013). Improving teaching and learning of basic maths and reading in Africa: Does teacher preparation count? *International Journal of Educational Development*, *33*(3), 272–282.

Atkinson, P. A. & Coffey, A. (2004). Analyzing documentary realities. In D. Silverman (Ed.), Qualitative research: Theory, method and practice (2nd ed.), London: Sage, 56–75.

Atteberry, A., & Bryk, A. S. (2011). Analyzing teacher engagement in literacy coaching activities. *Elementary School Journal*, 112(2), 356–382.

Avalos, B. (1998). School-based teacher development the experience of teacher professional groups in secondary schools in Chile. *Teaching and Teacher Education*, *14*(3), 257-271.

Bolam, R., McMahon, A., Stoll, L., Thomas, S., Wallace, M., Greenwood, A., ... & Smith, M. (2005). Creating and sustaining effective professional learning communities. *Bristol: University of Bristol y Departament of Education and Skills*.

Bloor, M. (1997). Techniques of validation in qualitative research: a critical commentary. *Context and method in qualitative research*.

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative research journal*, *9*(2), 27-40.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77-101.

Bryk, A., Camburn, E., & Louis, K. S. (1999). Professional community in Chicago elementary schools: Facilitating factors and organizational consequences. *Educational administration quarterly*, *35*(5), 751-781.

Buysse, V., Sparkman, K. L., & Wesley, P. W. (2003). Communities of practice: Connecting what we know with what we do. *Exceptional children*, *69*(3), 263-277.

Carey, J. W., & Gelaude, D. (2008). Systematic methods for collecting and analyzing multidisciplinary

team-based qualitative data. *Handbook for team-based qualitative research*, 227-274.

Corbin, J., & Strauss, A. (2008). Strategies for qualitative data analysis. *Basics of Qualitative Research. Techniques and procedures for developing grounded theory*, *3*.

Du Plessis, J., & Muzaffar, I. (2010). Professional Learning Communities in the Teachers' College: A Resource for Teacher Educators. *EQUIP1*.

DuFour, R. (2004). What is a" professional learning community"?. *Educational leadership*, *61*(8), 6-11.

Englert, C. S., & Tarrant, K. L. (1995). Creating collaborative cultures for educational change. Remedial and Special Education, 16(6), 325‑336, 353.

Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International journal of qualitative methods*, *5*(1), 80-92.

Fullan, M. G. (1991). *The new meaning of educational change*. New York, NY: Teacher College Press

Glaser, B. G. & Strauss, A. L. (1967). The discovery of grounded theory: Strategies for qualitative research. Chicago: Aldine

Geijsel, F. P., Sleegers, P. J. C., Stoel, R. D., & Krüger, M. L. (2009). The effect of teacher psychological and school organizational and leadership factors on teachers' professional learning in Dutch schools. *Elementary School Journal*, 109, 406 – 427.

Guest, G., MacQueen, K. M., & Namey, E. E. (2012). *Applied thematic analysis*. Sage Publications.

Hargreaves, A., & Fullan, M. G. (1992). *Understanding teacher development*. Teachers College Press, 1234 Amsterdam Avenue, New York, NY 10027.

Hipp, K. K., & Huffman, J. B. (2003). Professional Learning Communities: Assessment--Development--Effects.

Hollins, E. R., McIntyre, L. R., DeBose, C., Hollins, K. S., & Towner, A. (2004). Promoting a self-sustaining learning community: Investigating an internal model for teacher development. *International journal of qualitative studies in education*, *17*(2), 247-264.

Hord, S. (1997). *Professional learning communities: Communities of continuous inquiry and improvement.* Austin, TX: Southwest Educational Development Laboratory

Huffman, J. B., & Jacobson, A. L. (2003). Perceptions of professional learning communities. International

Journal of Leadership in Education: Theory and Practice, 6(3), 239‑250

Kruse, S. D., & Louis, K. S. (2007). Developing collective understanding over time: reflections on building professional community. *Professional Learning Communities: Divergence, Depth And Dilemmas: Divergence, Depth and Dilemmas*, 106.

Lee, J. C. K., Zhang, Z., & Yin, H. (2011). A multilevel analysis of the impact of a professional learning community, faculty trust in colleagues and collective efficacy on teacher commitment to students. *Teaching and teacher education*, *27*(5), 820-830.

Lee, D. H. L., & Lee, W. O. (2018). Transformational change in instruction with professional learning communities? The influence of teacher cultural disposition in high power distance contexts. *Journal of Educational Change*, *19*(4), 463-488.

Louis, K. S., & Marks, H. M. (1998). Does professional community affect the classroom? Teachers' work and student experiences in restructuring schools. *American Journal of Education*, 107, 532–575.

McLaughlin, M. W., & Talbert, J. E. (2001). *Professional communities and the work of high school teaching*. Chicago: University of Chicago.

Mitchell, C., & Sackney, L. (2000). *Profound improvement: Building capacity for a learning community.* Lisse, The Netherlands: Swets & Zeitlinger.

Newmann, F. M. (1996). *Authentic achievement: Restructuring schools for intellectual quality*. Jossey-Bass.

Newmann, F. M. (2006). The meaning of staff-shared understanding and commitment. *The keys to effective schools: Educational reform as continuous improvement*, 33.

Kruse, S. D., & Louis, K. S. (1993). An Emerging Framework for Analyzing School-Based Professional Community. Paper presented at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April 12-16, 1993).

Lincoln, Y. S. and Guba, E. G. (1985) Naturalistic Inquiry. Newbury Park, CA: SAGE.

Lomos, C., Hofman, R. H., & Bosker, R. J. (2011). Professional communities and student achievement–a meta-analysis. *School effectiveness and school improvement*, *22*(2), 121-148.

Louis, K. S., & Marks, H. M. (1998). Does professional community affect the classroom? Teachers' work and student experiences in restructuring schools. *American journal of education*, *106*(4), 532-575.

Philpott, C., & Oates, C. (2017). Professional learning communities as drivers of educational change: The case of learning rounds. *Journal of educational change*, *18*(2), 209-234.

Piper, B., Destefano, J., Kinyanjui, E. M., & Ong'ele, S. (2018). Scaling up successfully: Lessons from Kenya's Tusome national literacy program. *Journal of Educational Change*, *19*(3), 293-321.

Pryor, J. (1998). Action research in West African schools: problems and prospects. *International Journal of Educational Development*, *18*(3), 219-228.

Sackney, L., Walker, K., Mitchell, C., & Duncan, R. (2005). *Dimensions of school learning communities.* Paper presented at the annual conference of the International Congress for School Effectiveness and School Improvement, Barcelona, Spain.

Servage, L. (2008). Critical and transformative practices in professional learning communities. *Teacher education quarterly*, *35*(1), 63-77.

Sleegers, P., den Brok, P., Verbiest, E., Moolenaar, N. M., & Daly, A. J. (2013). Toward conceptual clarity: A multidimensional, multilevel model of professional learning communities in Dutch elementary schools. *The Elementary School Journal*, *114*(1), 118-137.

Stoll, L., Bolam, R., McMahon, A., Wallace, M., & Thomas, S. (2006). Professional learning communities: A review of the literature. *Journal of educational change*, *7*(4), 221-258.

Stoll, L., & Louis, K. S. (2007). *Professional learning communities: Divergence, depth and dilemmas*. McGraw-Hill Education (UK).

Strauss, A. & Corbin, J. (1990). Basics of qualitative research: Grounded theory procedures and techniques. Newbury Park, CA: Sage.

Strauss, A. & Corbin, J. (1998). Basics of qualitative research: Techniques and procedures for developing grounded theory (2nd ed.). London: Sage

Supovitz, J. A. (2002). Developing communities of instructional practice. Teachers College Record, 104(8), 1591‐1626

Tatto, M. T. (1999). Improving teacher education in rural Mexico: The challenges and tensions of constructivist reform. *Teaching and teacher education*, *15*(1), 15-35.

Toole, J., & Louis, K. S. (2002). The role of professional learning communities in international education. In K. Leithwood & P. Hallinger (Eds.), *Second international handbook of educational leadership and administration* (pp. 245–281). Dordrecht, The Netherlands: Kluwer.

Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and teacher education*, *24*(1), 80-91.

Zhang, J., & Pang, N. S. K. (2016). Exploring the characteristics of professional learning communities in China: A mixed-method study. *The Asia-Pacific Education Researcher*, *25*(1), 11-21.

Appendix 6.A. Essential characteristics-dimensions of Professional Learning Communities highlighted by the literature

	Shared values, norms, vision	Focus on student learning - Collective responsibility	Reflective dialogue - reflective professional inquiry	Deprivatizing practice - shared personal practice	Collaboration - collective learning and application	Normative control	Group, as well as individual, learning is promoted	Mutual trust	Inclusive membership	Openness, networks and partnerships	Socialization of new professional members	Shared and supportive leadership	Physical conditions and Human Capacities
Kruise & Louis (1993)	x	x	x	X	x								
Newmann et al. (1996)	x	x	x	X	x								
Hord (1997)	x			X	x							x	x
Bryk et al (1999)			x	X	x	x					x		
Bolam et al (2005)	x	x	x		x		x	x	x	x			
Stoll et al. (2006)	x	x	x		x		x	x	x	x			

Appendix 6.B. Document Analysis

Country	*Project/ Organization*	*Documents reviewed*
Democratic Republic of Congo	Opportunities for Equitable Access to Quality Basic Education (OPEQ) / International Rescue Committee (IRC)	• White paper: Teacher Learning Circle Case Study: (Post)Crisis Katanga Province, DRC; IRC (2014). • Frisoli, P. (2014). Teachers' experiences of professional development in (post)crisis Katanga province, Southeastern Democratic Republic of Congo: A case study of teacher learning circles (dissertation), University of Massachusetts. • Final Report on the Impact of the OPEQ Intervention in the Democratic Republic of Congo. (draft) NYU and IRC, (2016).
Democratic Republic of Congo	Accelerating Access and Learning in the DRC (ACCELERE!) Project / Chemonics International	• The ACCELERE! Continuing Professional Development Program (FHI 360 program document).
Equatorial Guinea	Program for Educational Development of Equatorial Guinea (PRODEGE) / FHI 360	• Teacher Learning Circle Guides for Areas 1 – 4, 16 guides (2018). • Módulo de Formación Docente para la Acreditación Docente: Competencias de la docencia transformadora en preescolar y primaria (Teacher's module for Diplomado in-service teacher training program.) (2018). • ProFADs technical paper (paper that describes structure of PRODEGE teacher development programs) (2017).
Ghana	USAID Partnership for Education: Learning / FHI 360	• USAID Partnership for Education: Learning: School-based INSET Guide (2017). • USAID Partnership for Education: Learning: Coaching/School-based INSET Resource Guide. (2017). • USAID Partnership for Education: Learning: Coaching/School-based INSET Facilitator Guide (2017).
Nigeria	Reading and Numeracy Activity (RANA) / FHI 360	• Proposal for FHI 360's expanded scope of work in Nigeria, GEP 3/ EAC, UNICEF Nigeria (2017) • RANA Final Report (2018) • Meeting log form for RANA teachers
Senegal	Passerelles / FHI 360	• Facilitator Guide
South Africa	Department of Basic Education / VVOB	• Professional Learning Communities: A guideline for South African schools. (2015). https://southafrica.vvob.be/sites/southafrica/files/professional_learning_communities_a_guideline_for_south_african_schools_0_0.pdf
Tanzania	WEKEZA project / International Rescue Committee (IRC)	• Trainer's toolkit: Learning to Read in a Healing Classroom • Teacher Learning Circles (TLCs) Meeting Notes format • Organizing and Implementing Teacher Learning Circles: School-based toolkit
Other	International Network for Education in Emergencies (INEE)	• The Facilitator's Guide • Peer Coaching Toolkit, Level 1 - TLCs only • Peer Coaching Toolkit, Level 2 - TLCs plus classroom observations https://inee.org/resources/teachers-crisis-contexts-peer-coaching-pack

Appendix 6.C. Guidelines for semi-structured interviews

Table M.1: Individuals interviewed by project and role

Country / Project with PLC component	**Individuals interviewed**
Ghana / Learning	Chief of Party, Ghana Learning
Southern Senegal / USAID PASSERELLES, Nigeria / Addressing Education in the North East Nigeria (AENN)	Associate Director, Education in Conflict and Crisis
Equatorial Guinea / Program for Educational Development of Equatorial Guinea (PRODEGE)	Technical Advisor of Active Learning, PRODEGE
Nigeria / RANA	Technical Advisor, Early Grade Learning

Interview Guidelines

The idea of the school as a professional learning community is relatively new and we are investigating how it is applied in developing countries. The definition we use in this research is:

"a group of education professionals sharing and critically interrogating their practice in an ongoing, reflective, collaborative, inclusive, learning-oriented, growth-promoting way, with the collective purpose of enhancing student learning**"** (adapted from Stoll et al., 2006)

1) Which specific nomenclature does your project use to refer to PLCs?
2) What is the PLC definition used in this project?
3) What is the main purpose of PLCs in this project?
4) For how long have PLCs been implemented under this project?
5) In this project, what were the main facilitators to:

(a) establishing PLCs?

(b) sustaining PLCs?

6) In this project, what were the main barriers to:

(a) establishing PLCs?

(b) sustaining PLCs?

7) Which specific classroom practices does the program expect teachers to change as a result of their participation in PLCs?
8) What are the hypothesized channels through which the project expects PLCs to change the teaching practices highlighted above?
9) What specific areas are PLCs in your project designed to support?
10) How is the PLCs meeting frequency determined?
11) How frequently are PLCs expected to meet?
12) How long each PLC meeting should last?
13) At what level are PLCs in this project organized?
14) At what educational level are PLCs promoted?
15) How are PLCs organized?
16) Are teachers assigned to participate or do they volunteer?
17) Who participate in the PLCs?

18) Does the project establish a minimum and a maximum number of PLC participants?
19) Please indicate which materials were developed and provided by the project to PLCs.
20) What is the main purpose-objective of each of these materials?
21) To what extent to materials provided guide the meetings?
22) Do the PLCs in this project have a facilitator?
23) If yes, who plays the role of a facilitator?
24) What are the main tasks the facilitator is responsible for?
25) Does participation in the PLC have any professional implications for teachers?
26) Do teachers receive any incentive to attend PLCs?
27) Does the project promote participation of external experts in the PLCs?
28) If so, what is the objective of having an external expert participate?
29) Does the project promote the analysis of student work or student assessment data under the PLCs?
30) If yes, please specify the type of student work or assessment data analyzed within the PLCs:
31) Are PLCs part of a broader professional development intervention?
32) If so, please indicate the purpose of the PLCs within the broader professional development program.
33) For each of the following statements, indicate if your project:

1= does not recognize this practice under the PLC design or promotes it

2= recognizes this practice under the PLC design but does not consistently promote it

3= explicitly recognize and consistently promote this PLC practice

De-privatization of Practice	
PLCs provide opportunity for teachers to observe one another in their practice and offer feedback.	1 2 3
PLCs provide opportunity for teachers to have open and honest conversations about their teaching practices.	1 2 3
Reflective dialogue	
PLCs provide opportunity for teachers to have deep discussions and reflection around specific teaching practices.	1 2 3
PLCs provide opportunity for teachers to have deep discussions and analysis of student work.	1 2 3
Focus on student learning	
PLCs focus on what and how well students are learning rather than how teachers are teaching.	1 2 3

PLCs focus on not simply ensuring that students are taught but to ensure that they learn.	1 2 3
PLCs provide opportunity for teachers to analyze student work or student data.	1 2 3
Collaboration	
PLCs promote collaboration that go beyond superficial exchanges, to incorporate a deep discussion of practice.	1 2 3
PLCs provide opportunity for teachers to collaboratively review student work as a way to share and improve instructional practices.	1 2 3
Shared norms and values	
PLC members agree on shared values about children and their ability to learn	1 2 3
PLC members agree on shared values around the roles of teachers	1 2 3
PLC members agree shared norms about the use of time and space within the school	1 2 3

Appendix 6.D. Expert validation sessions

Date	*Expert(s) working on the following project*	*Number of participants*
6/7/2017	Guatemala / MCC-funded Improving the Quality of Secondary Education project	1
6/13/2017	Equatorial Guinea / PRODEGE	2
6/16/2017	Nigeria /RANA	1
6/19/2017	Nigeria /RANA	2
6/21/2017	Ghana / Learning	1
11/6/2018	PLCs in general and across several FHI 360 and IRC projects	1
1/28/2019	El Salvador/USAID Education of Children and Youth Project	1

Chapter 7. Toward a Deeper Understanding: Testing a Multidimensional Framework of Professional Learning Communities in Sub-Saharan African Schools[35]

7.1. Introduction

Governments and donors are increasingly funding in-service training programs aimed at improving teaching quality to ensure that all children in school obtain the skills and knowledge they are meant to acquire (Cilliers et al., 2018). Despite increasing efforts to improve the quality of teaching and a growing body of research demonstrating that teachers are among the most important determinants of student learning, "teacher training programs vary enormously, both in their form and their effectiveness" (Popova et al., 2018, p. 4). Evidence from low- and middle-income countries suggests that short teacher trainings, usually held in a central location, are not effective in improving teacher practice (Popova et al., 2018). This is often the case in Sub-Saharan Africa, where teacher training is delivered through short courses centrally planned by the Ministry of Education on specific topics, such as the introduction of new curricula (Lauwerier & Akkari, 2015), and are not connected to the context or the on-the-ground reality teachers experience in their school communities. Possible reasons for the failure of teacher training programs in developing countries may rely on their focus on imparting knowledge, on being overly theoretical, and on their delivery being too rote and passive (Loyalka, forthcoming). All these characteristics make it difficult for these training programs to be translated into practical pedagogical improvements as teaching is a skill that needs to be developed through ongoing practice (Kennedy 2016).

In an increasing recognition that teachers' knowledge is situated in teachers' daily experiences and best learned through ongoing practice and reflection, many countries, including developing countries, have recently resorted to alternative models of teacher professional development, such as Professional Learning Communities, or PLCs. PLCs emerged in the past two decades as part of a paradigm shift, which replaces traditional and isolated forms of teacher professional development with a more connected approach embedded in the context (Darling-Hammond & Richardson, 2009; Du Plessis & Muzaffar, 2010; Vescio, Ross & Adams, 2007). PLCs attempt to move professional development beyond the acquisition of new knowledge and skills (Vescio, Ross & Adams, 2007) to the reflection and improvement of classroom practice through collaboration among teachers (DuFour, 2004). In this sense, PLCs reflect a shift away

[35] This is based on: Soares F and Galisson K. Toward a Deeper Understanding: Testing a Multidimensional Typology of Professional Learning Communities in Sub-Saharan African Schools. This project is supported by FHI 360.

from traditional approaches of professional development which emphasize "knowledge for practice" to an approach grounded on the idea of "knowledge of practice" (Cochran-smith & Lytle, 1999).

Although PLCs are in vogue in many developing countries, the concept of PLC originated and has been studied and more widely applied in the context of developed countries, mainly the United States, Netherlands and England. To the best of our knowledge two main systematic reviews have been conducted on the effect of PLCs on teaching practices, none of which include studies from low- and middle-income countries. Cordingley and colleagues (2005) examine the effect of collaborative Continuing Professional Development (CPD[36]) on both teaching and learning and do not restrict their review to PLCs. The review concluded that collaborative CPD can impact a number of teacher outcomes, including: greater confidence; enhanced beliefs in teachers' power to make a difference in pupils' learning; enthusiasm for collaborative working; greater commitment to changing practice; and enhanced knowledge and practice. Vescio, Ross and Adams (2008) reviewed studies assessing the effect of PLCs on teaching practices. The authors found that all 11 research articles reviewed supported the idea that well-developed PLCs can have a positive effect on teaching practices. Most of the evidence covered by both reviews is based on qualitative or quantitative cross-sectional data (Vescio, Ross & Adams, 2008), and experimental (or quasi-experimental) and/or long-term designs are nearly absent in the literature[37]. There is only limited research on PLCs in the context of developing countries (exceptions include Tatto, 1999; Pryor, 1998; Avalos, 1998; Abrahams, 1997), which is mostly qualitative in nature and does not explore how conceptual models and frameworks of PLCs drawing from the Western literature may or may not be applicable in the context of developing countries.

This study improves our understanding of the specific characteristics and experiences of PLCs in the context of Sub-Saharan African countries and explore to what extent the conceptual foundations, models and empirical findings from the Western literature are also applicable to the reality of developing countries. PLCs in developing countries may operate in different ways reflecting their unique context, society and culture. PLCs in developed countries, for example, emphasize teachers' autonomy and authority in making decisions regarding the processes, agenda and objective of their learning communities. Previous research, however, has indicated that developing countries trying to improve very low-functioning education systems should focus on introducing highly specific approaches to instructional change (Piper et al., 2018). So, while teachers' autonomy is a key element of PLCs in developed countries, developing countries may adopt more prescriptive approaches to promoting effective instructional change. This might be needed to ensure

[36] The authors defined Collaborative CPD as teachers working with at least one other related professional on a sustained basis.
[37] Exceptions include Supovitz (2002) and Dunne et al. (2000).

PLCs are not reinforcing and maintaining existing traditional substandard practice rather than changing it (Stoll et al., 2006).

This research builds off the framework proposed in chapter six, which integrates elements of PLCs found in the literature from developed countries with specific features more commonly found in Sub-Saharan African countries identified through a qualitative research. We first examine the extent to which higher levels of participation in PLCs are associated with higher constructivist and student-centered teaching using Ordinary Least Square (OLS) regression. This model includes school-fixed effects to account for unobserved differences between schools and controls for teacher background characteristics and number of months since the PLC has been established. Constructivist and student-centered teaching is measured through teachers' self-reported classroom practices. Second, we test the relationship between the strength of a PLC using a PLC Strength Index and constructivist and student-centered teaching practices also employing OLS regression and the same control variables as the previous model. We measure PLC strength using teacher's self-reporting of the presence of specific PLC characteristics informed by the PLC framework presented in chapter 6 and the Western literature. Because the PLC Strength Index is a variable at the school level, we are unable to include school-fixed effects, but we employ cluster-robust standard errors. Third, we expand this model to include three structural varying characteristics of the PLC framework and explore how these features interact with PLC strength to influence constructivist student-centered teaching. Overall, because of the endogeneity issues we are unable to address through our estimation strategies, we cannot attribute causality. Rather, the objective of our research is to document empirically the associations between participation in PLCs and teaching outcomes, as well as between core and varying characteristics of PLCs and teaching outcomes.

Data for this study was collected from three countries - Equatorial Guinea, Ghana and Nigeria, reaching a sample of over 2,300 preschool and primary education teachers. The three countries were selected because at the time of the data collection they had ongoing, large-scale, teacher professional development programs with a strong emphasis on the development of PLCs. The sampling method and total sample reached varied by country. Overall, given the sampling methods adopted, the final sample reached is not statistically representative of the overall teacher population in each country.

The research contributes to the literature in three main ways. First, this is the first study in the context of developing countries to examine the relationship between participation in PLCs and teaching practices. Second, we also test the relationship between the strength of a PLC and teaching practices. To our knowledge, only one study by Louis and Marks (1998) has analyzed the relationship between classroom practices and the existence of PLC core characteristics that inform its strength. Additionally, with this analysis we test the extent to which PLC core characteristics highlighted in the Western literature that

inform the strength of a PLC are also valid in developing countries. Third, we explore the mechanisms through which the strength of a PLC may influence constructivist student-centered teaching practices. To the best of our knowledge, this is the first effort to quantitatively explore the mechanisms through which PLCs may influence teaching practices.

7.2. A Typology and Framework of PLCs

Research on PLCs is still in the early stages of theory building (Sleegers et al., 2013) and there is no broad consensus in the literature on a definition of PLC (Lomos, 2012; Vescio, Ross, & Adams, 2007; Stoll et al., 2006). However, most definitions agree that PLCs involve a group of teachers sharing and critically interrogating their practice in an ongoing, reflective, collaborative, inclusive, learning-oriented, growth-promoting way (Stoll and Louis, 2007; Stoll et al., 2006; Toole and Louis, 2002; Mitchell and Sackney, 2000;).

In chapter 6 of this thesis we conducted a document review of PLCs as well as semi-structured interviews with practitioners to inform a PLC typology that accounts for how PLCs in Sub-Saharan African countries operate. We reviewed project and policy documents that detailed the design of a PLC, with a specific focus on PLCs at the primary education level. The typology categorizes PLCs into three models - autonomous, structured and scripted - and accounts for specific features that vary greatly across PLC types. This typology of PLCs was than integrated with dimensions previously proposed by the Western literature to form one cohesive conceptual framework. The framework that emerged from the study is organized around five interrelated PLC dimensions: core characteristics, model, structural features, supportive conditions and stages of maturity. The dimensions of core characteristics, supportive conditions and stages of maturity emerged directly from the literature from developed countries (Atteberry & Bryk, 2011; Geijsel et al., 2009; Stoll et al., 2006; Bolam et al., 2005; McLaughlin & Talbert, 2001; Mitchell & Sackney, 2000; Bryk et al., 1999; Hord, 1997; Newmann et al., 1996; Kruise & Louis, 1993; Fullan, 1991), while the type of model and structural features were added by the authors based on a inductive content analysis of documents and interview transcripts. The purpose of the framework is to improve conceptual clarity around the multidimensional nature of PLCs, integrate elements that might be specific to low- and middle-income countries and guide policy-makers and practitioners in low- and middle-income countries on the design and promotion of PLCs.

7.3. Research setting

This study was carried out in three Sub-Saharan African countries – Equatorial Guinea, Nigeria and Ghana - in 2018. These countries were selected for the research as they have large-scale teacher professional development programs with a strong emphasis on the development of PLCs. Additionally, we sought to

sample countries in which different models of PLC are employed and that at least two models of PLCs – structured and scripted – are present, as they are more commonly used in Sub-Saharan African countries.

Equatorial Guinea

Equatorial Guinea's education system faces enormous challenges. Half of primary students are overage for their grade, 24% of primary students repeat a grade, and only 22% of students eligible for secondary school ultimately enroll (UNICEF, 2016). To address these challenges, the Program for Educational Development of Equatorial Guinea (PRODEGE), funded by Kosmos Energy and Trident Energy and the Government of Equatorial Guinea, focuses on expanding teacher professional development, strengthening the capacity of local and national institutions, and fully integrating active learning into preschool, primary, and lower secondary schools. Specifically, PRODEGE implements a nation-wide, two-year teacher professional development program for 1,500 preschool and primary school teachers who lack teaching credentials. In addition, all primary and preschool teachers in the country (nearly 8,000 teachers) are encouraged to participate in a cluster-based PLC, called Teacher Learning Circles. As a source of peer support, the Circles provide teachers the opportunity to reflect and learn from each other's experiences in the classroom. PRODEGE adopts a *scripted* PLC model; PLCs have scripted guides developed by the program to guide the content and pace of each PLC meeting. Trained coordinators facilitate a bi-monthly four-hour meeting by going through the exercises in the guides. Teachers who participate in PLC meetings and nation-wide training during the two academic years of implementation receive a teacher upgrading certificate.

Nigeria

Nigeria has 510 living indigenous languages. Although this linguistic diversity represents a rich cultural heritage, it also poses challenges for the education system as the Nigerian Federal Ministry of Education has mandated that local languages be used for instruction in the early grades. Research has found that teachers are underprepared for mother-tongue instruction and curriculum materials for mother-tongue instruction are not supplied in schools (Duze, 2011). To help overcome these challenges, the Reading and Numeracy Activity (RANA) project, funded by the U.K. Department for International Development (DFID) through the United Nations Children's Fund (UNICEF), aims to improve the quality of literacy and numeracy instruction for children in primary grades 1–3 in approximately 200 schools in the two states (Katsina and Zamfara), with the ultimate goal of improving learning outcomes.

To achieve these goals, RANA has developed Hausa-language teaching and learning materials and promotes an integrated approach to teachers' professional development that encompasses hands-on workshops, ongoing coaching to teachers and peer learning through school-based PLCs, called school-based communities of practice. School-based PLCs in Nigeria encourage teachers to meet on a weekly basis

for about an hour. The PLC adopts a *structured* model and follows a sequenced approach to each meeting: opening the meeting with a letter sound movement or song; discussion of a weekly success; discussion of a weekly challenge; practice of one instructional skill; practice lesson for upcoming week; and finally feedback on the practice lesson. The meetings are facilitated by lead teachers, or high performing teachers who have received a few hours of training on how to run the meetings following the *structured* approach.

Ghana

Low learning outcomes in early grades is an area of concern in Ghana. According to the 2013 and 2015 Early Grade Reading Assessment conducted by the National Education Assessment Unit (Ghana Education Service, NEAU 2013 and 2015), only the top 2% of children in primary grade 2 could read with fluency and comprehension. To address some of these challenges, the Ministry of Education with support from USAID launched the Partnership for Education: *Learning*, a five-year program that aims to improve performance in reading for students in kindergarten 2 (second and last year of kindergarten before pupils enter primary education), primary grades 1 and 2.

The *Learning* program promotes teacher professional development to upgrade teaching practices in early grade literacy. The multi-faceted program promotes teacher training in new methodologies, provides scripted materials for teaching and guides for students all in the mother tongue language, provides coaching to teachers and encourages teachers' participation in school-based INSET (SBI) meetings to improve their practice. The SBI meetings are a form of PLC at the school-level that have been institutionalized by the Ministry of Education but have not been consistently utilized by teachers and directors. *Learning* promotes the use of the SBI meeting space as peer learning opportunities for teachers to reflect on how their reading classes are going, get support to troubleshoot challenges, and learn about reading and instruction in more depth than could be covered in the face-to-face workshops. Starting in 2017, the project provided kindergarten and early primary grade teachers with scripted materials to guide the one-hour weekly meetings which are facilitated by Head Teachers and Curriculum Leads. Following our typology, this PLC is an example of the *scripted* model.

Figure 7.1. Comparative PLC design by country

	Equatorial Guinea	Ghana	Nigeria
Model	Scripted	Scripted	Structured
Type of grouping	Cluster-based	School- based	School- based
Meeting materials	Highly scripted guides	Highly scripted guides	Meeting log
Facilitation	Trained facilitator	No trained facilitator	Trained facilitator
Minimum encouraged meeting frequency*	Twice per month	Weekly	Weekly
Grade Level	Preschool and primary	Kindergarten 2, Primary grades 1 & 2	Primary grades 1 & 2

** While this refers to the minimum frequency encouraged by the program, teachers may meet more often than that based on the agreement between teachers within a PLC on the desired frequency.*

7.4. Data

We use data from a Teacher Survey administered to preschool and primary school teachers in Equatorial Guinea, Nigeria and Ghana between March and July of 2018. The survey was implemented by Family Health International 360 (FHI 360) in partnership with the Ministry of Education and project teams in each of the three countries. The purpose of the survey was to gather information on teachers' background characteristics, teachers' instructional beliefs and practices, frequency of participation in PLCs, and PLCs' core characteristics and structural features. The Teacher Survey was designed to be self-administered and was filled out completed independently by the teacher without the need for an interviewer.

The specific sampling method and total sample reached varied by country, but overall the survey adopted a nonprobability sampling approach, *i.e.*, participant teachers were not randomly selected from the teacher population. This means that the sampling techniques employed were not intended to be used to infer from the sample to the general population in statistical terms. Hence, we cannot claim that the final sample in each country is representative of the teacher population in that country.

Table 7.1 describes the final sample size reached in each country as well as teachers' key background characteristic. In Nigeria, the survey targeted all grades 1 and 2 teachers within the 200 schools in two states where the RANA project operates. The survey was applied by trained project staff after a school cluster meeting, which convenes teachers from four to six schools on a monthly basis. Out of approximately 800 grades 1 and 2 teachers that participate in RANA, 618 participated in the survey. In Equatorial Guinea, the Teacher Survey was administered in 17 training sites spread across the country where PRODEGE delivers an in-service certification program to approximately 1,500 preschool and primary teachers. In total, 1,349 teachers responded to the survey. In Ghana, the survey was carried out in a sample of 86 schools, in 10 regions where the *Learning* project is implemented. The survey was administered by Monitoring and

Evaluation Officers during regular monitoring trips to schools. A total of 355 teachers in preschool, grades 1 and 2 responded to the survey.

Table 7.1. Sample size and teachers' background characteristic by country

	Ghana			Nigeria			Equatorial Guinea		
	Mean	SD	Obs	Mean	SD	Obs	Mean	SD	Obs
Years of Experience	12.28	6.00	344	10.90	5.94	567	9.37	4.89	1291
Female	0.89	0.31	349	0.16	0.37	593	0.78	0.42	1349
SES Index	0.77	0.21	350	0.60	0.28	613	0.56	0.32	1344
Permanent Contract	0.97	0.17	350	0.85	0.36	603	0.56	0.50	1349
Observations			355			618			1349

7.5. Methodology

Index construction and variables used in the analysis

The main dependent variable of interest is *constructivist student-centered teaching*. This is a composite measure, which aggregates teachers' self-rating on their constructivist beliefs about learning and instruction as well as on their self-reported student-centered classroom practices. The items that form this measure come from the OECD Teaching and Learning International Survey (TALIS). To develop the index, we conducted exploratory factor analyses to identify one meaningful higher-order construct and reduce the number of items to the most theoretically and empirically relevant underlying construct. The variable is standardized with a mean of 0 and a standard deviation of 1.

The study explores two main independent variables. *High Participation in PLCs* is a binary variable that takes the value of 1 if the teacher participated in PLC meetings three or more times in the past four weeks and 0 otherwise. Although the variable also takes the value of 0 if the teacher did not attend a PLC meeting in the past four weeks, only a small percentage of teachers (less than 2% in all three countries) reported not attending a PLC meeting. As there is no theoretical or empirical recommendation in the literature as to how often teachers should participate in PLC meetings, our definition of high participation in PLCs is derived from interviews with practitioners and document reviews on PLC design. The consultations indicated that teachers are expected to meet at minimum twice per month in order for participation to have a meaningful impact on teaching practices.

The second main independent variable, *PLC Strength Index*, is a composite index. In constructing the index, we first measured each PLC core characteristic by calculating their raw score on Likert-scales adapted from

Supovitz (2002) and Louis and Mark (1998). For each core characteristic scale, teachers reported the extent to which they agreed on a number of PLC-related items ranging from 1 = "strongly disagree" to 5 = "strongly agree", the frequency with which they have implemented specific practices, ranging from 1= "never" to 6="once a week or more", or how much time they dedicate to specific topics, ranging from 1= "less than 25% of the entire meeting time" to 4= "more than 85% of entire meeting time". Table 7.2 presents illustrative items for each scale while Appendix 7.D presents the full questionnaire with all complete scales. For each PLC core characteristic, we generated an index that equals 1 if the average score in the scale was high and 0 otherwise.

We then conducted an exploratory factor analysis, which revealed that four out of the five core characteristics measure one overarching construct of PLC. EFA results illustrated in Table 7.3 show that only one factor emerged with an eigenvalue greater than 1. Factor loading for each core characteristic was 0.3 or higher, with the exception of shared purpose and focus on learning in Nigeria and de-privatized practice in all three countries[38]. While the four characteristics seem to be measuring the internal strength of a PLC, de-privatized practice might be capturing teacher interactions (co-teaching, observation, etc.) that happen outside of the PLC meetings. We construct the Index using the four core characteristics that load well, and we analyze de-privatized practice separately. The final PLC Strength Index ranges from 0 to 4. Since in Ghana and Nigeria PLCs are at the school level, for these two countries we generated an average PLC Strength Index value for each school. These results represent an important contribution to an ongoing debate in the literature on whether PLC core characteristics are separate factors or one unidimensional construct **(**Sleegers et al., 2013).

Table 7.2. Sample items from each PLC core characteristic scale used in the construction of the PLC Strength Index

PLC core characteristic Scale	Sample Item
Shared norms and values	My colleagues in the PLC share my values about children and their ability to learn.
Collaboration	The discussions in my PLC go beyond superficial exchanges and incorporate a deep discussion of practice.
De-privatized practice	How often have you **v**isited other PLC colleagues' classrooms to observe and discuss their teaching?
Reflective Dialogue	In a typical PLC meeting, about how much time is spent on analyzing teaching practices? Teachers discuss specific teaching practices of team members.
Focus on student learning	In my PLC we collaboratively review student work and student data as a way to improve instructional practices.

Note: Full survey instrument and complete scales provided in Appendix 7.D.

[38] Only one study by Bryk and colleagues (1999) used factor analyses to determine whether the five characteristics of PLCs measure a single underlying construct. Consistent with our results, they found that de-privatized practice showed lower factor loading.

Table 7.3. Factor 1 Loadings for PLC core characteristics

	Nigeria	Ghana	E.G.
Shared Norms and Values	0.4193	0.2557	0.3832
Focus on Learning	0.5131	0.2396	0.3280
Reflective Dialogue	0.6404	0.6814	0.5591
Collaborative	0.7337	0.7366	0.6116
De-privatized Practice	-0.1096	0.0701	0.1984

* *Factor 1 eigenvalue was 1.39958 for Nigeria; 1.13456 for Ghana and 0.98043 for E.G.*

We also include in our analysis secondary independent variables of interest, which are structural varying characteristics of PLCs that are directly linked to the type of PLC model defined in the typology. Since Equatorial Guinea and Ghana adopt a scripted model of PLCs, we include the variable *guided-manual*, which equals to 1 if the content of the PLC meeting is determined by the guided manual and 0 otherwise. While we would expect all PLCs to use the guided-manual provided by the programs, as per the descriptive statistics discussed below, the material take-up was relatively low. For Nigeria, which adopts a structured model of PLC, we include the variable *content jointly*, which equals to 1 if the content of the PLC meeting is determined jointly and autonomously by the PLC members and 0 otherwise. For all three countries, we include the variable *active facilitator*, which is also a binary variable that equals to 1 if the facilitator encourages discussion and promotes dialogue and 0 otherwise.

Table 7.4 presents descriptive statistics on key variables of interest. Full descriptive statistics are presented in Appendix 7.A. Although we cannot do a direct comparison of means as we have not tested for cross-cultural comparability or "invariance" of the indices we use, the descriptive statistics are helpful in understanding teaching practices and PLC strength within each of the countries. Total raw scores for *constructivist student-centered teaching* are presented in Table 7.4 for easiness of interpretation. The mean score was similar across all three countries and relatively high, ranging from 3.29 in Nigeria to 3.46 in Ghana (out of a 4.5 maximum score). On average, 74% of teachers exhibit *high PLC participation* in Ghana, compared to 67% in Nigeria and 47% in Equatorial Guinea. The lower participation found in Equatorial Guinea is not surprising, as teachers have to commute to a central meeting place to participate in the cluster-based PLC. *PLC Strength Index* was relatively low in all three countries: in a five-point index, we registered 1.67 in Ghana, 2.43 in Nigeria and 2.17 in Equatorial Guinea. In both Ghana and Nigeria approximately 60% of the PLCs have an active facilitator, while in Equatorial Guinea this percentage is much higher at 78%. This is also expected as in Equatorial Guinea FHI 360 provided training in PLC facilitation techniques to almost all PLC coordinators. In both Ghana and Equatorial Guinea, which adopt a scripted model of PLC, 94% of teachers reported having received the guided-manual distributed by the

program, while only about 30% of teachers reported using the guided-manual to inform the content of their meetings. An analysis of this variable at the school level for Ghana and Nigeria show that for the most part all teachers within the same PLC agree whether guided-manual is used to guide the discussions or not. In Nigeria, the content for most PLC meetings (85%) was determined jointly by the PLC members, which is expected under the structured PLC model adopted in the country.

Table 7.4. Descriptive statistics on key variables

	Ghana			Nigeria			E.G.		
	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.
Dependent Variables									
Constructivist student-centered teaching raw score	344	3.46	0.48	589	3.29	0.51	1254	3.43	0.46
Main Independent Variables									
High Participation	338	0.74	0.44	553	0.67	0.47	1255	0.47	0.50
PLC Strength Index	355	1.67	1.24	618	2.43	1.27	1349	2.17	1.34
Secondary Independent Variables									
Active Facilitator	339	0.61	0.49	574	0.60	0.49	1273	0.78	0.42
Guided-manual	305	0.31	0.47				1166	0.28	0.45
Content Jointly				460	0.85	0.35			

Note: Full descriptive statistics are provided in the Appendix 7.A.

Estimation Strategy

We estimate the influence of high participation in PLCs on teaching beliefs and practices using an Ordinary Least Square (OLS) regression model, controlling for teacher background characteristics and time the PLC has been in place. From an empirical perspective, determining the effect of participation in PLCs on teaching beliefs and practices is challenging because of omitted variable bias and self-selection bias and the resulting endogeneity issues. Because teachers who participate in PLCs may be different in observable and unobservable ways from those who do not participate or participate less frequently, the relationship between PLC participation and teaching practices detected through a traditional OLS estimation may be capturing unobservable influences on teaching practices that are omitted as predictors. Unfortunately, because no information is available on number of teachers per school in this dataset, we cannot employ the same Instrumental Variable approach as we did in Chapter 4 of this dissertation. However, while we cannot account for the unobservable differences, we notice that teachers who exhibit high participation in PLCs are very similar on key observable socio-economic background characteristics compared to those who show low participation (see Appendix 7.A. for descriptive statistics based on level of participation). Using Coarsened Exact Matching we matched teachers on observable characteristics and the results (available in Appendix 7.C.) remain very similar to the results from the main specification we propose.

Another potential source of bias are unobserved differences between schools that may influence teaching outcomes. We address this by including school fixed-effects in our main estimation. In fixed-effects models, the school effect is no longer part of the residual. Instead, a dummy variable is included for each school, which controls for all variation (observed and unobserved) at the school level. Hence, the model identifies the effect of PLC participation on teaching outcomes *within* schools rather than *between* schools. School fixed-effects also account for the attrition of teachers to schools. Because we do not have data on which school teachers in Equatorial Guinea belong to, the fixed-effect model is included only for Ghana and Nigeria. In Equatorial Guinea, we use region fixed-effects instead and results have to be interpreted more carefully, as we are not able control for within school variation.

We define the fixed-effects ordinary least square model as:

$$Yis = \beta 0 + \beta 1\, Xis + \beta 2 Wis + \cdots + \beta 1 + Wri + \varepsilon_{is} \quad (1)$$

where Y_{is} is the teaching outcome for teacher *i* in school *s* ; X_i is the ith observation of the binary explanatory variable high participation in PLCs; W_{is},….,W_{ri} are the ith observation of each of the control variables (covariates); u_s is school dummy variable coefficient and ε_i is the error term. The model controls for teacher background characteristics (years of experience, gender, university degree, socio-economic status index) as well as for the number of months since the PLC has been established.

We estimate the association between PLC Strength Index and teaching beliefs and practices also using an OLS regression model and the same control variables as the previous model. Because the PLC Strength Index is a variable at the school level, we are unable to include school-fixed effects. We do, however, employ cluster-robust standard errors for Nigeria and Ghana that corrects the standard errors for within-cluster (schools) error correlation. Since we do not have data on which school teachers in Equatorial Guinea belong to, we employ robust standard errors but are unable to adjust for clustering by school.

We define the cluster robust ordinary least square model as:

$$Yis = \beta 0 + \beta 1\, Zis + \beta 2 Wis + \cdots + \beta 1 + rWri + \varepsilon_{is} \quad (2)$$

where Y_{is} is the teaching outcome for teacher *i* in school *s* ; Z_i is the ith observation of the explanatory variable PLC Strength Index; W_{is},….,W_{ri} are the ith observation of each of the control variables (covariates); u_s is school dummy variable coefficient and ε_i is the error term.

In a second stage, we also add to this model structural varying characteristics of PLCs that are directly linked to the type of PLC model defined in the typology: guided-manual, content jointly, and active facilitator. We interact each of structural varying characteristics with the PLC Strength Index to understand

how the relationship between these variables and the strength of a PLC influence our key teaching outcomes.

Overall, because of the endogeneity issues described in this section that we are unable to address through our estimation strategies, we cannot attribute causality and hence cannot claim that participation in PLCs or strong PLCs cause certain instructional practices or beliefs. The objective of our research is to document empirically the associations between participation in PLCs and teaching outcomes, as well as between core and varying characteristics of PLCs and teaching outcomes. We seek to demonstrate that strong PLCs or high participation are present when these instructional practices-beliefs occur.

7.6. Results

Relationship between high participation in PLCs and teaching beliefs and practices

Table 7.5 presents the results of the OLS estimates, examining the relationship between high participation in PLCs and constructivist student-centered teaching, with school fixed-effects for Ghana and Nigeria. Full regression tables are presented in Appendix 7.C. The results show that teachers in Nigeria and Ghana who participate more frequently in PLCs are significantly more likely to report higher constructivist student-centered teaching. On average, teachers who exhibit high participation in PLCs are associated with an increase of approximately 0.392 standard deviations in the constructivist student-centered teaching score in Ghana and 0.316 standard deviations in Nigeria. Appendix 7.B. presents a kernel distribution of raw as well as standardized scores on the constructivist student-centered teaching scale. It should also be noted that Equatorial Guinea was the only country where we were unable to include school-fixed effects in the regression analysis. Unobserved and observed school factors influencing the teaching outcome of interest that we were unable to control for could have affected our results. A sensitivity analysis (Table 7.C.4 Appendix 7.C.) using a different cut-off point for the variable high participation (participation in PLCs two or more times per month) yield insignificant results, indicating that our findings are sensitive with respect to the construction of the variable. Our results suggest that one or two meetings a month may not be sufficient to drive associations with constructivist student-centered teaching and that higher levels of participation in PLCs may be needed to have an influence on the teaching outcomes of interest.

Our results may also indicate that teachers' interaction in PLCs was working to change existing traditional teaching practices towards more active pedagogical approaches. Bryk and et al. (1999) suggest that "if a professional community in fact fosters instructional change, it does so by creating an environment that supports learning through innovation and experimentation" (p. 771). This is an important finding in light of the debate in the literature on whether PLCs promote reformation or transformation of practice (Lee et al., 2018; Philpott & Oates, 2017; Servage, 2008). Reformation of practice involves aligning individual

teaching practices with prescribed norms or existing mandated goals (Philpott & Oates, 2017), whereas transformation of practice involves questioning goals and empowering teachers to be "… producers of pedagogical knowledge rather than solely consumers and implementers" (Philpott & Oates, 2017, p. 213). The association we find may suggest a transformation of practice towards more student-centered approaches in contexts characterized by high teacher-centered instruction.

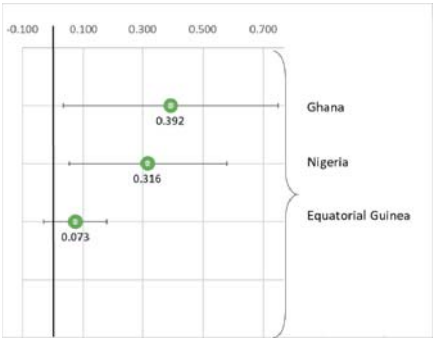
These results also shed light on previous mixed findings of quantitative literature investigating the relationship between elementary school teachers' participation in PLCs and teaching outcomes. Sargent (2014) tests the relationship between frequency of participation in school-level PLCs and teachers' reported use of innovative pedagogy. She finds that teachers who participate more frequently in PLC activities in the school are more likely to report frequent use of innovative teaching methods. Supovitz (2002) compared team-based (a form of PLC*)* and non-team-based teachers on their instructional practices. Contrary to the findings of Sargent (2014), he does not find any difference between elementary school teachers in team-based and non-team-based schools on their individual instructional practices (Supovitz, 2002).

Table 7.5. Regression results: coefficients of constructivist student-centered teaching associated with PLC High Participation, by country

	GHANA	**NIGERIA**	**EQUATORIAL GUINEA**
	Constructivist student-centered teaching	Constructivist student-centered teaching	Constructivist student-centered teaching
PLC High Participation	0.392*	0.316*	0.073
	(0.218)	(0.160)	(0.063)
Fixed-effects	Yes	Yes	No
Observations	303	422	1070

Notes: (i) school fixed effects included for Nigeria and Ghana and region fixed-effects for Equatorial Guinea.
(ii) *** indicates statistical significance at 1%, ** at 5% and * at 10%. Full regression tables provided in Appendix 7.C.

Figure 7.2. Relationship between high participation on PLCs with constructivist student-centered teaching, by country



Note: For Equatorial Guinea the model specification does not account for school fixed-effects, only region fixed-effects are included. Hence, the three effect sizes are not completely comparable.

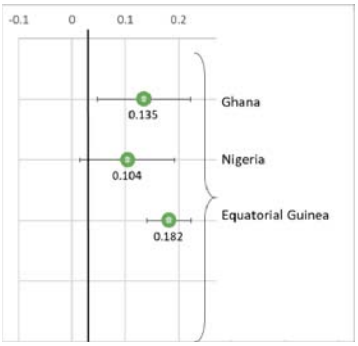
The analysis presented in Table 7.6 examines the relationship between PLC Strength Index and constructivist student-centered teaching. The results are striking and demonstrate that across all three countries, stronger PLCs are significantly associated with higher constructivist student-centered teaching. Each point in the PLC Strength Index (which means adding one core PLC characteristic) is associated with an increase of 0.118 standard deviations in constructivist student-centered teaching in Ghana, 0.112 in Nigeria and 0.192 in Equatorial Guinea. This is in line with earlier quantitative literature by Louis and Marks (1998) analyzing the relationship between the quality of classroom pedagogy and the existence of the core characteristics of PLCs. The authors find that to the extent the five core characteristics of PLCs are present, the higher the social support for student achievement and authentic pedagogy.

Table 7.6. Estimation of PLC strength index on constructivist student-centered teaching

	GHANA	**NIGERIA**	**EQUATORIAL GUINEA**
	Constructivist student-centered teaching	Constructivist student-centered teaching	Constructivist student-centered teaching
PLC Strength Index	0.135**	0.104*	0.182***
	(0.053)	(0.054)	(0.025)
SE clustered by schools	Yes	Yes	No
Observations	303	420	1070

Notes: (i) robust standard errors are adjusted for clustering by school for Nigeria and Ghana.
(ii) *** indicates statistical significance at 1%, ** at 5% and * at 10%. Full regression tables provided in Appendix 7.C.

Figure 7.3. Influence of PLC Strength Index on constructivist student-centered teaching, by country



Note: For Equatorial Guinea the model specification includes robust standard errors, but they are not adjusted for clustering by school, as in Ghana and Nigeria. Also, the PlC Strength Index in Equatorial Guinea is not a school level average. Hence, the three effect sizes are not completely comparable.

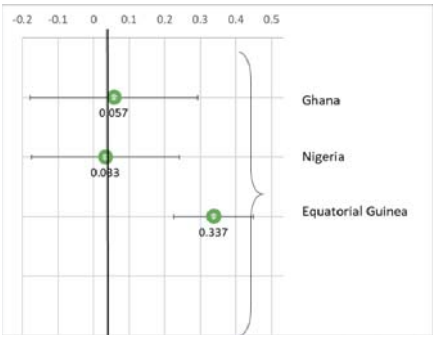
We also analyze the relationship between de-privatized practice (a core PLC characteristic that did not load well into the PLC Strength Index) and constructivist student-centered teaching. Results illustrated in Table 7.7 show insignificant relationships for Ghana and Nigeria, but for Equatorial Guinea the presence of high de-privatized practices is significantly associated with a 0.337 standards deviation increase in constructivist student-centered teaching. Since the model specification for Equatorial Guinea does not adjust standard errors for clustering by school, as in Ghana and Nigeria, the effect sizes are not completely comparable. In addition, de-privatized practice in Equatorial Guinea is not a school level average, as it is for Ghana and Nigeria. While PRODEGE implemented PLCs only in its last couple years in Equatorial Guinea, since the very beginning the 10-year program promoted feedback, suggestions and observations between teachers as a strategy to incentivize student-centered teaching. This ongoing practice was likely reflected in the dynamic of the PLCs once they started and may indicate that for meaningful and effective de-privatized practices to happen and to have an effect on teaching, more time is needed.

Table 7.7. Estimation of De-privatized practice on constructivist student-centered teaching

	GHANA Constructivist student-centered teaching	**NIGERIA** Constructivist student-centered teaching	**EQUATORIAL GUINEA** Constructivist student-centered teaching
De-privatized practice	0.057	0.033	0.337***
	(0.143)	(0.126)	(0.068)
SE clustered by schools	Yes	Yes	No
Observations	300	420	1051

Notes: (i) robust standard errors are adjusted for clustering by school for Nigeria and Ghana. (ii) *** indicates statistical significance at 1%, ** at 5% and * at 10%. Full regression tables provided in Appendix 7.C.

Figure 7.4. Influence of De-privatized practice on constructivist student-centered teaching



Note: For Equatorial Guinea the model specification includes robust standard errors, but they are not adjusted for clustering by school, as in Ghana and Nigeria. In addition, De-privatized in Equatorial Guinea is not a school level average. Hence, the three effect sizes are not completely comparable.

Table 7.8 expands the model proposed in Table 5 to explore the mechanisms through which the strength of a PLC may influence constructivist student-centered teaching. We test the association between constructivist student-centered teaching and the interaction between key structural varying features directly linked to the PLC models from the typology with the PLC strength index. Although a few coefficients from Table 7 are not significant (which may be due to a lack of statistical power given our small sample size and clustered analysis), the direction and magnitude of the coefficients are helpful in describing and explaining the relationships among key variables of interest.

We observe from Table 7.8 that in Ghana and in Nigeria having an active facilitator in the PLC meetings is positively and significantly correlated to *constructivist student-centered teaching*. However, the interaction between *active facilitator* and PLC strength index is negative in both cases, indicating that this relationship varies according to the different levels of PLC strength. For low, and in some cases medium values of the PLC Strength Index, having an active facilitator is positively associated with constructivist student-centered teaching. However, the higher the PLC strength, the lower the benefit of having a facilitator on *constructivist student-centered teaching*, and at some point of the PLC strength index this relationship actually becomes negative. Overall, the higher the PLC strength, the lower the benefit of having an active facilitator in these two countries. In Equatorial Guinea this relationship is not significant and in terms of magnitude is close to 0.

We also observe from Table 7.8 that *guided-manual* is positively associated to *constructivist student-centered teaching* in both countries that adopt them, Ghana and Equatorial Guinea, although insignificant

in the case of Ghana. As with *active facilitator*, the interaction between *PLC strength* and *guided-manual* is negative, which indicates that having the content of a meeting determined by a guided-manual is positively associated to constructivist student-centered teaching for low and medium values of the PLC strength index, and negatively associated with this outcome for high values of the index. In the specific case of Nigeria, we find that having the content of the meetings be determined jointly by PLC members is negatively associated with constructivist student-centered teaching, although this relationship is not significant. The coefficient for the interaction between *content jointly* and *PLC strength* is positive, indicating that the relationship between *content jointly* and *constructivist student-centered teaching* is negative for low and medium values of the PLC strength index and positive otherwise.

Table 7.8. Estimation of PLC Strength Index and varying features on teaching beliefs and practices

	GHANA Constructivist student-centered teaching	**NIGERIA** Constructivist student-centered teaching	**EQUATORIAL GUINEA** Constructivist student-centered teaching
PLC Strength Index	0.272***	0.078	0.187***
	(0.094)	(0.102)	(0.062)
Active Facilitator	0.318*	0.467**	-0.050
	(0.176)	(0.211)	(0.128)
Guided Manual	0.064		0.205*
	(0.173)		(0.122)
Content Jointly		-0.341	
		(0.215)	
PLC Strength Index * Active Facilitator	-0.151	-0.160*	0.017
	(0.107)	(0.092)	(0.064)
PLC Strength Index * Guide Manual	-0.084		-0.100*
	(0.144)		(0.056)
PLC Strength Index * Content Jointly		0.138	
		(0.110)	
SE clustered by schools	Yes	Yes	No
Observations	273	339	908

Notes: (i) robust standard errors are adjusted for clustering by school for Nigeria and Ghana. (ii) *** indicates statistical significance at 1%, ** at 5% and * at 10%.

7.7. Conclusion and Indicative Policy Implications

Our results show a positive significant relationship between higher levels of participation in PLC and constructivist student-centered practices for two out of the three countries in the study. When exploring the relationship between the strength of a PLC and teaching practices, we find striking results across the three countries indicating that stronger PLCs are associated with higher constructivist student-centered practices. Since we control for number of months since the PLC has been established, it is unlikely that this influence is due to level of maturity of the PLC. For a fifth core PLC characteristic analyzed separately, de-privatized practice, we only found evidence of association with constructivist student-centered practices in Equatorial Guinea. Descriptive statistics exploring mechanisms show that the lower the PLC strength, the higher the benefits of having an active facilitator and a guided manual on constructivist student-centered practices. Finally, in the specific case of Nigeria, having the content of the meeting be determined jointly by PLC members in weak or medium strength PLCs is negatively associated with our teaching outcomes of interest.

We acknowledge important limitations of this study. First, we cannot attribute causality as our study is merely exploratory. Instead, our goal is to document empirically the linkages between our variables of interest, and we cannot say that PLCs causes constructivist student-centered practices, but we do seek to demonstrate that these practices occur when PLCs are present. More longitudinal as well as experimental or quasi-experimental research on PLCs is warranted. Second, we rely on teachers' self-reported use of constructivist student-centered practices in the classroom. Even though the survey data was kept confidential and it was clearly explained to teachers that no incentive would result from the survey, unobserved characteristics, such ambition or ability, could be related with the likelihood of over reporting. Third, teachers' unobservable characteristics might co-determine how they answer items about teaching practices and how they perceive the characteristics of their PLC. For instance, teachers more open to innovate pedagogical approaches may more frequently apply student-centered practices and may have a different view of their PLC, one that favors deeper collaboration and focus on student learning. Future research that relies on observational data of teaching practices and of PLC strength is recommended to fill out this gap.

Despite the limitations, we believe this research has important indicative policy implications. First, our results indicate that higher participation in PLCs may be beneficial in promoting the use of innovative classroom instructional practices. More importantly, it suggests that scripted PLCs can potentially lead to transformation of practice rather than working as a vehicle for implementation of authority-endorsed change or mandates (Lee & Lee, 2018). Second, the results underline the importance of policies and programs that aim to promote PLCs to create a common vision focused on fostering at least four out of five core characteristics from the PLC framework in conjunction, not separately. This finding also indicates that these

four core characteristics conceptualized by the Western literature may be applicable in the context low-and middle-income countries. There is a need, however, for research focusing more attention on how these four core characteristics can emerge and be sustained. Third, our results suggest that a scripted model may be helpful in supporting the work of PLCs that are just starting or are weak through the development of materials that can help stimulate and guide teachers' work. On the contrary, full autonomous models for PLCs that are still weak may have a negative influence on teaching outcomes. Finally, training facilitators on active facilitation techniques as PLCs start may be beneficial for influencing teaching practices

References

Abrahams, M. (1997). Collaborative space in South African schools: A comparative perspective. Journal of Negro Education, 66(4), 409-422.

Altinyelken, H. K. (2010). Pedagogical renewal in sub- Saharan Africa: the case of Uganda. *Comparative Education*, *46*(2), 151–171.

Akyeampong, K., Lussier, K., Pryor, J., & Westbrook, J. (2013). Improving teaching and learning of basic maths and reading in Africa: Does teacher preparation count? *International Journal of Educational Development*, *33*(3), 272–282.

Atteberry, A., & Bryk, A. S. (2011). Analyzing teacher engagement in literacy coaching activities. *Elementary School Journal*, 112(2), 356–382.Avalos, B. (1998). School-based teacher development the experience of teacher professional groups in secondary schools in Chile. *Teaching and Teacher Education*, *14*(3), 257-271.

Bolam, R., McMahon, A., Stoll, L., Thomas, S., Wallace, M., Greenwood, A., ... & Smith, M. (2005). Creating and sustaining effective professional learning communities. *Bristol: University of Bristol y Departament of Education and Skills*.

Bryk, A., Camburn, E., & Louis, K. S. (1999). Professional community in Chicago elementary schools: Facilitating factors and organizational consequences. *Educational administration quarterly*, *35*(5), 751-781.

Cilliers, J., Fleisch, B., Prinsloo, C., & Taylor, S. (2018). How to improve teaching practice? Experimental comparison of centralized training and in-classroom coaching. *Research on Improving Systems of Education (RISE): Working Paper.* (RISE-WP-18-024).

Cochran-Smith, M., & Lytle, S. L. (1999). Chapter 8: Relationships of knowledge and practice: Teacher learning in communities. *Review of research in education*, *24*(1), 249-305.

Cordingley, P., Bell, M., Thomason, S., & Firth, A. (2005). The impact of collaborative continuing professional development (CPD) on classroom teaching and learning. *Review: How do collaborative and sustained CPD and sustained but not collaborative CPD affect teaching and learning*.

Darling-Hammond, L., & Richardson, N. (2009). Research review/teacher learning: What matters. *Educational leadership*, *66*(5), 46-53.

Du Plessis, J., & Muzaffar, I. (2010). Professional Learning Communities in the Teachers' College: A Resource for Teacher Educators. *EQUIP1*.

DuFour, R. (2004). What is a" professional learning community"?. *Educational leadership*, *61*(8), 6-11.

Duze, C. O. (2011). Implementation of the Mother-Tongue/Language Component of the National Policy on Education in Nigeria. *Lwati: A Journal of Contemporary Research, 8* (1).

Geijsel, F. P., Sleegers, P. J. C., Stoel, R. D., & Krüger, M. L. (2009). The effect of teacher psychological and school organizational and leadership factors on teachers' professional learning in Dutch schools. *Elementary School Journal*, 109, 406 – 427.

Ghana Education Service (2014). Ghana 2013 Early Grade Reading Assessment and Early Grade Mathematics Assessment: Report of Findings.

Ghana Education Service, RTI International, & Education Assessment and Research Centre (2016). Ghana 2015 Early Grade Reading Assessment and Early Grade Mathematics Assessment: Report of Findings.

Hipp, K. A. & Huffman, J. B. (2003). *Professional Learning Communities: Assessment— Development— Effects*. Paper presented at the Annual Conference of the International Congress for School Effectiveness and Improvement. Sydney, Australia, January.

Hord, S. (1997). *Professional learning communities: Communities of continuous inquiry and improvement.* Austin, TX: Southwest Educational Development Laboratory

Hollins, E. R., McIntyre, L. R., DeBose, C., Hollins, K. S., & Towner, A. (2004). Promoting a self-sustaining learning community: Investigating an internal model for teacher development. *International journal of qualitative studies in education*, *17*(2), 247-264.

Kennedy, M. M. (2016). How does professional development improve teaching?. *Review of Educational Research* 86(4), 945–980.

Kruse, S. D., & Louis, K. S. (1993). An Emerging Framework for Analyzing School-Based Professional Community. Paper presented at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April 12-16, 1993).

Lauwerier, T., & Akkari, A. (2015). Teachers and the quality of basic education in sub-Saharan Africa. *Education Research and Foresight: Working Papers Series*, (11).

Little, J. W. (1990). The persistence of privacy: Autonomy and initiative in teachers' professional

relations. *Teachers college record*, *91*(4), 509-536.

Little, J. W. (2003). Inside teacher community: Representations of classroom practice. *Teachers college record*, *105*(6), 913-945.

Lomos, C., Hofman, R. H., & Bosker, R. J. (2011). Professional communities and student achievement–a meta-analysis. *School effectiveness and school improvement*, *22*(2), 121-148.

Louis, K. S., & Marks, H. M. (1998). Does professional community affect the classroom? Teachers' work and student experiences in restructuring schools. *American journal of education*, *106*(4), 532-575.

Loyalka, P., Popova, A., Li, G., & Shi, Z. (Forthcoming). Does Teacher Training Actually Work? Evidence from a Large-Scale Randomized Evaluation of a National Teacher Training Program. *American Economic Journal: Applied Economics*.

Mitchell, C., & Sackney, L. (2000). *Profound improvement: Building capacity for a learning community.* Lisse, The Netherlands: Swets & Zeitlinger

Newmann, F. M. (1996). *Authentic achievement: Restructuring schools for intellectual quality*. Jossey-Bass.

Newmann, F.M. (2006). The meaning of staff-shared understanding and commitment. In W.D. Hawley., and D.L. Rollie (Eds.), *The keys to effective schools*. Thousand Oaks: Corwin Press

OECD (2009). Creating Effective Teaching and Learning Environments: First Results from TALIS. Paris: OECD. Retrieved August 07, 2009 from http://www.oecd.org/dataoecd/17/51/43023606.pdf

Popova, A., Evans, D. K., & Arancibia, V. (2016). Inside in-service teacher training: What works and how do we measure it?. In *RISE annual conference 2016*.

Popova, A., Evans, D. K., Breeding, M. E., & Arancibia, V. (2018). Teacher Professional Development around the World: The Gap between Evidence and Practice. *World Bank group: Policy Research Working Paper*, (8572).

Pryor, J. (1998). Action research in West African schools: problems and prospects. *International Journal of Educational Development*, *18*(3), 219-228.

Sargent, T. C. (2014). Professional learning communities and the diffusion of pedagogical innovation in the Chinese education system. *Comparative Education Review*, *59*(1), 102-132.

Sleegers, P., den Brok, P., Verbiest, E., Moolenaar, N. M., & Daly, A. J. (2013). Toward conceptual clarity: A multidimensional, multilevel model of professional learning communities in Dutch elementary

schools. *The Elementary School Journal*, *114*(1), 118-137.

Stoll, L., Bolam, R., McMahon, A., Wallace, M., & Thomas, S. (2006). Professional learning communities: A review of the literature. *Journal of educational change*, *7*(4), 221-258.

Strahan, D. (2003). Promoting a collaborative professional culture in three elementary schools that have beaten the odds. *The Elementary School Journal*, *104*(2), 127-146.

Supovitz, J. A. (2002). Developing communities of instructional practice. *Teachers College Record*, *104*(8), 1591-1626.

Tatto, M. T. (1999). Improving teacher education in rural Mexico: The challenges and tensions of constructivist reform. *Teaching and teacher education*, *15*(1), 15-35.

Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and teacher education*, *24*(1), 80-91.

World Bank (2018). *World Development Report 2018 - LEARNING to Realize Education's Promise. Washington, D.C.*

Appendix 7.A. Descriptive Statistics

Table 7.A.1. PLC descriptive statistics

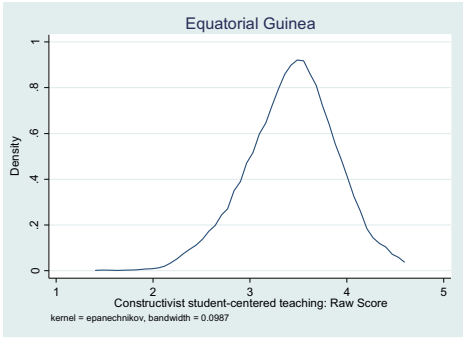
	Ghana			Nigeria			Equatorial Guinea		
	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.
Dependent Variable									
Constructivist student-centered teaching	344	3.46	0.48	589	3.29	0.51	1254	3.43	0.46
Independent Variables									
High Participation	338	0.74	0.44	553	0.67	0.47	1255	0.47	0.50
PLC Strength Index	355	1.17	1.11	618	2.19	1.25	1349	1.89	1.22
PLC Strengh Index Items									
Shared Purpose	340	4.06	0.55	563	4.34	0.54	1322	4.18	0.52
Collaborative Activity	338	3.00	0.64	557	3.36	0.77	1265	3.46	0.69
Focus Learning	337	3.56	1.16	544	3.86	1.04	1210	4.17	0.78
Reflective Dialogue	335	2.20	0.85	549	2.88	0.82	1252	2.78	0.86
Deprivatized practice	333	3.99	1.20	554	3.55	1.13	1296	3.35	1.24
PLC varying features									
PLC Duration	337	0.83	0.38	568	0.80	0.40	1297	0.83	0.38
PLC Establishment	323	5.10	3.45	535	10.79	7.52	1264	8.47	7.21
# Teachers	339	6.93	3.16	568	5.77	3.43	1292	15.99	5.01
Facilitator	334	0.68	0.47	573	0.98	0.15	1272	1.00	0.07
Active Facilitator	339	0.61	0.49	574	0.60	0.49	1273	0.78	0.42
Manual	339	0.94	0.23				1304	0.98	0.13
Guide Manual	305	0.31	0.47	460	0.07	0.26	1166	0.28	0.45
Content Jointly	305	0.66	0.47	460	0.85	0.35	1166	0.61	0.49
Expert	329	0.43	0.50	567	0.53	0.50	1270	0.76	0.43
Observations	355			618			1349		

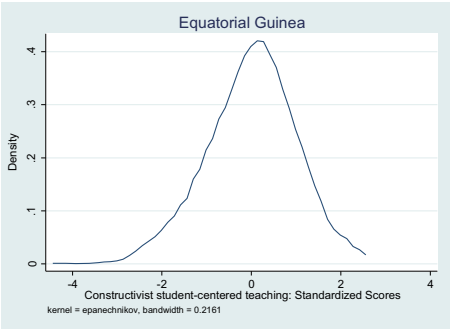
Table 7.A.2. Teachers' Descriptive statistics by level of participation in PLCs

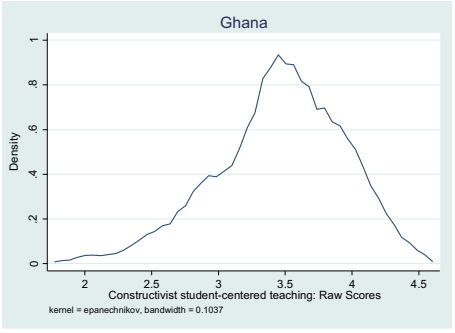
	Obs	Low Participation (L)	High Participation (H)	Diff(H-L)
Ghana				
Experience	330	12.28	12.29	0.01
Female	336	0.91	0.84	-0.07*
University	338	0.60	0.60	-0.00
SES Index	337	0.77	0.74	-0.03
Permanent Contract	337	0.97	0.97	-0.01
Nigeria				
Experience	510	10.89	11.21	0.32
Female	534	0.19	0.10	-0.09**
University	544	0.51	0.42	-0.09**
SES Index	550	0.61	0.59	-0.02
Permanent Contract	541	0.89	0.84	-0.06*
Equatorial Guinea				
Experience	1201	9.45	9.24	-0.20
Female	1255	0.79	0.78	-0.00
University	1240	0.40	0.41	0.01
SES Index	1250	0.58	0.54	-0.04**
Permanent Contract	1255	0.53	0.57	0.04

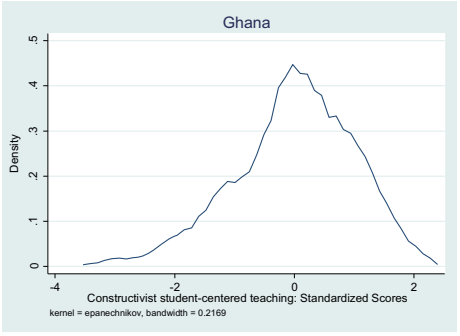
Notes: Significance is denoted as: * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

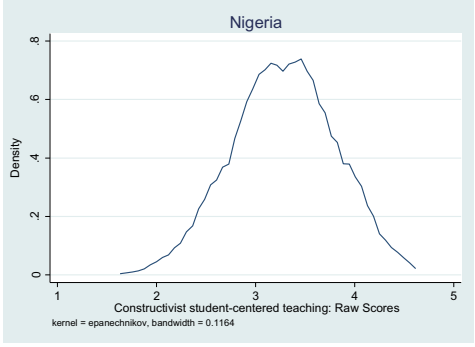
Appendix 7.B. Kernel Density distribution of dependent variable constructivist student-centered teaching

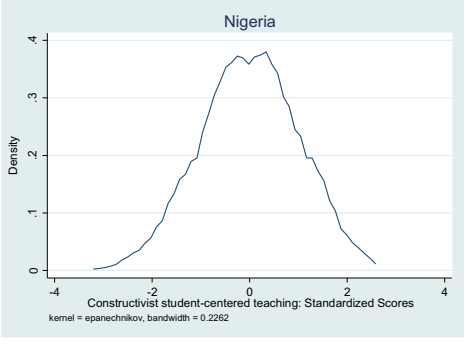












Appendix 7.C. Results Tables

Table 7.C.3. Estimation of teachers' high participation in PLCs on constructivist student-centered teaching

	GHANA Constructivist student-centered teaching	**NIGERIA** Constructivist student-centered teaching	**EQUATORIAL GUINEA** Constructivist student-centered teaching
High Participation	0.392*	0.316*	0.073
	(0.218)	(0.160)	(0.063)
Experience	-0.021*	-0.011	0.012
	(0.011)	(0.010)	(0.008)
Female	-0.238	0.260	-0.098
	(0.205)	(0.161)	(0.075)
University	0.136	0.213*	0.044
	(0.139)	(0.121)	(0.035)
SES Index	-0.013	-0.151	0.061
	(0.331)	(0.227)	(0.102)
Permanent Contract	0.013	-0.067	-0.025
	(0.370)	(0.238)	(0.075)
PLC Establishment	0.023	-0.005	-0.007*
	(0.021)	(0.010)	(0.004)
Constant	0.210	0.110	-0.407
	(0.857)	(0.945)	(0.348)
Observations	303	422	1070

Notes: (i) school fixed effects included for Nigeria and Ghana and region fixed-effects for Equatorial Guinea.
(ii) *** indicates statistical significance at 1%, ** at 5% and * at 10%.

Table 7.C.4: Sensitivity analysis: Estimation of teachers' high participation in PLCs on constructivist student-centered teaching using a different cut-off point for high participation

	GHANA Constructivist student-centered teaching	**NIGERIA** Constructivist student-centered teaching	**EQUATORIAL GUINEA** Constructivist student-centered teaching
High Participation	0.596	0.118	0.208
	(0.362)	(0.245)	(0.157)
Experience	-0.019*	-0.009	0.012
	(0.011)	(0.010)	(0.008)
Female	-0.240	0.260	-0.097
	(0.205)	(0.163)	(0.075)
University	0.151	0.211*	0.040
	(0.139)	(0.122)	(0.035)
SES Index	-0.075	-0.151	0.063
	(0.332)	(0.229)	(0.102)
Permanent Contract	0.016	-0.058	-0.032
	(0.370)	(0.240)	(0.074)
PLC Establishment	0.026	0.000	-0.008*
	(0.021)	(0.010)	(0.004)
Constant	0.018	0.434	-0.534
	(0.901)	(0.975)	(0.362)
Observations	303	422	1070

Notes: (i) school fixed effects included for Nigeria and Ghana and region fixed-effects for Equatorial Guinea.
(ii) *** indicates statistical significance at 1%, ** at 5% and * at 10%.

Table 7.C.5: Estimation of teachers' high participation in PLCs on constructivist student-centered teaching employing Coarsened Exact Matching

	GHANA Constructivist student-centered teaching	**NIGERIA** Constructivist student-centered teaching	**EQUATORIAL GUINEA** Constructivist student-centered teaching
High Participation	0.496	0.306**	0.047
	(0.308)	(0.154)	(0.066)
Experience	-0.020*	-0.005	0.011
	(0.011)	(0.013)	(0.008)
Female	-0.389	0.023	-0.126
	(0.358)	(0.209)	(0.081)
University	0.142	0.241*	0.056
	(0.144)	(0.136)	(0.045)
SES Index	0.043	-0.163	0.046
	(0.387)	(0.235)	(0.105)
Permanent Contract	0.022	-0.093	0.029
	(0.256)	(0.245)	(0.077)
PLC Establishment	0.014	-0.005	-0.009*
	(0.030)	(0.010)	(0.004)
Constant	0.210	0.135	-0.469
	(0.718)	(0.309)	(0.438)
Observations	287	390	1,046

Notes: (i) school fixed effects included for Nigeria and Ghana and region fixed-effects for Equatorial Guinea.
(ii) *** indicates statistical significance at 1%, ** at 5% and * at 10%.

Table 7.C.6. Estimation of PLC Strength Index on constructivist student-centered teaching

	GHANA Constructivist student-centered teaching	**NIGERIA** Constructivist student-centered teaching	**EQUATORIAL GUINEA** Constructivist student-centered teaching
PLC Strength Index	0.135**	0.104*	0.182***
	(0.053)	(0.054)	(0.025)
High Participation	0.210	0.078	0.033
	(0.127)	(0.107)	(0.062)
Experience	-0.020**	-0.012	0.011
	(0.009)	(0.010)	(0.008)
Female	-0.325*	0.293*	-0.087
	(0.166)	(0.154)	(0.073)
University	0.306**	0.145	0.046
	(0.130)	(0.100)	(0.034)
SES Index	0.013	-0.052	0.057
	(0.264)	(0.191)	(0.100)
Permanent Contract	0.231	-0.066	-0.054
	(0.199)	(0.165)	(0.073)
PLC Establishment	0.019	-0.008	-0.006
	(0.020)	(0.007)	(0.004)
Constant	-0.440	-0.108	-0.710**
	(0.335)	(0.227)	(0.349)
Observations	303	420	1070

Notes: (i) robust standard errors are adjusted for clustering by school for Nigeria and Ghana. (ii) *** indicates statistical significance at 1%, ** at 5% and * at 10%.

Table 7.C.7. Estimation of De-privatized practice on constructivist student-centered teaching

	GHANA Constructivist student-centered teaching	**NIGERIA** Constructivist student-centered teaching	**EQUATORIAL GUINEA** Constructivist student-centered teaching
De-privatized practice	0.057	0.033	0.337***
	(0.143)	(0.126)	(0.068)
High Participation	0.213	0.077	0.051
	(0.136)	(0.107)	(0.063)
Experience	-0.019**	-0.015	0.014*
	(0.009)	(0.010)	(0.008)
Female	-0.266	0.317**	-0.094
	(0.163)	(0.158)	(0.075)
University	0.288**	0.137	0.045
	(0.134)	(0.101)	(0.036)
SES Index	-0.028	-0.004	0.027
	(0.255)	(0.193)	(0.100)
Permanent Contract	0.249	-0.086	-0.015
	(0.202)	(0.165)	(0.075)
PLC Establishment	0.022	-0.008	-0.008*
	(0.021)	(0.007)	(0.004)
Constant	-0.337	0.112	-0.520
	(0.393)	(0.198)	(0.359)
Observations	300	420	1051

Notes: (i) robust standard errors are adjusted for clustering by school for Nigeria and Ghana. (ii) *** indicates statistical significance at 1%, ** at 5% and * at 10%.

Appendix 7.D. Survey Instrument

Section 1: Background Information

DIRECTIONS: Read each question. Carefully check the one answer that fits best.

1. Region and District where you currently work:

2a. Name of school where you currently work:

2b. School EMIS code:

3. Are you female or male?

Male	☐
Female	☐

4. Which grades do you currently teach at this school? (*check all that apply*)

Preschool 1	☐
Preschool 2	☐
Primary 1	☐
Primary 2	☐
Primary 3	☐
Primary 4	☐
Primary 5	☐
Primary 6	☐

5. What is the highest level of formal education you have completed?

ISCED level 1	☐
ISCED level 2	☐
ISCED level 3	☐
ISCED level 4	☐
ISCED level 5B	☐
ISCED level 5A	☐
ISCED level 6	☐

6. Did you complete a teacher education program?

Yes	☐
No	☐

7. How many years of experience do you have working as a teacher at this school?

1 2 3 4 5 6 7 8 9

10 11 12 13 14 15 16 17 18 19

☐ 20 or more

8. How many years of experience do you have working as a teacher in total?

1 2 3 4 5 6 7 8 9

10 11 12 13 14 15 16 17 18 19

☐ 20 or more

9. What is your employment status as a teacher at this school?

Permanent employee	☐
Fixed-term contract for a period of more than 1 school year	☐
Fixed-term contract for a period of 1 school year or less	☐

DIRECTIONS: For questions 10-16 please choose yes or no.
Does your household have:

		Yes	No
10	electricity?	☐	☐
11	a television?	☐	☐
12	a refrigerator?	☐	☐

Does any member of your household own:

		Yes	No
13	a cell phone?	☐	☐
14	a computer?	☐	☐
15	a bicycle?	☐	☐
16	a car or truck?	☐	☐

Section 2: General teaching

17. Please indicate to what extent you agree or disagree with the following statements about teaching:

	Strongly agree	Agree	Not sure	Disagree	Strongly Disagree
Effective/good teachers demonstrate the correct way to solve a problem.	□	□	□	□	□
Instruction should be built around problems with clear, correct answers, and around ideas that most students can grasp quickly.	□	□	□	□	□
How much students learn depends on how much background knowledge they have; that is why teaching facts is so necessary.	□	□	□	□	□
A quiet classroom is generally needed for effective learning.	□	□	□	□	□
My role as a teacher is to facilitate students' own inquiry.	□	□	□	□	□
Students learn best by finding solutions to problems on their own.	□	□	□	□	□
Students should be allowed to think of solutions to practical problems themselves before the teacher shows them how they are solved.	□	□	□	□	□
Thinking and reasoning processes are more important than specific curriculum content	□	□	□	□	□

18. How often does each of the following happen throughout the school year? Please mark one choice in each row.

	Never or almost never	Occasionally	Frequently	In all or nearly all lessons
I present a summary of recently learned content.	□	□	□	□
Students work in small groups to come up with a joint solution to a problem or task.	□	□	□	□
I give different work to the students who have difficulties learning and/or to those who can advance faster.	□	□	□	□
I refer to a problem from everyday life or work to demonstrate why new knowledge is useful.	□	□	□	□
I let students practice similar tasks until I know that every student has understood the subject matter.	□	□	□	□

19. Please indicate to what extent you agree or disagree with the following statements about teaching:

	Strongly agree	Agree	Not sure	Disagree	Strongly Disagree
I think that teamwork is important.	□	□	□	□	□
People who work in teams can learn more than if they work by themselves.	□	□	□	□	□
I think I can get better results if I work in cooperation with others rather than individually.	□	□	□	□	□

Section 3: Professional Learning Community (PLC) meetings

20. Do you currently participate in a PLC meetings?

Yes	□
No	□

IF YOU DO NOT CURRENTLY PARTICIPATE IN A PLC, PLEASE STOP THIS SURVEY HERE. WE THANK YOU FOR YOUR PARTICIPATION!

21. For how many months has your PLC been in place and actively meeting?

0 1 2 3 4 5 6 7 8 9
) ! @ # $ % ^ & * (

□ 20 or more

22. How many teachers typically participate in your PLC meeting?

0 1 2 3 4 5 6 7 8 9
) ! @ # $ % ^ & * (

□ 20 or more

23. Which of the following statements is true about the size of your PLC meeting?

The group size is adequate for a participatory discussion and reflection.	□
Discussion is only possible if the group breaks up into smaller groups.	□
The size is not conducive to a participatory discussion and reflection.	□

24. How often does your PLC meet?

Twice per week or more	□
Once per week	□
Twice per month	□
Once per month	□
Every other month	□
Once per semester	□
Once per year	□

25. How many PLC meetings were held in the last 4 weeks?

0 1 2 3 4 5 6 7 8

□ 09 or more

26. How many PLC meetings have you attended in the last four weeks?

0 1 2 3 4 5 6 7 8

□ 09 or more

27a. On average, how long does each PLC meeting usually last?

30 minutes or less	□
30 to 60 minutes	□
60 to 90 minutes	□
90 to 120 minutes	□
121 minutes or more	□

27b. Which of the following statements is true about the length of time of your PLC meetings?

We usually have enough meeting time to cover the topics the group would like to cover.	□
We usually don't have enough meeting time to cover all the topics we would like to cover.	□

28a. Does your PLC have a designated facilitator?

Yes	□
No	□

28b. If so, who is your PLC meeting facilitator?

School director	□
Teacher	□
Other	□

28c. If you have a designated facilitator, which statement(s) are true about her/his participation in the PLC meeting? *(mark all that apply)*

The facilitator ensures that our PLC meetings take place.	□
The facilitator guides the group discussion.	□
The facilitator encourages group members to share experiences with one another during discussions.	□
The facilitator takes attendance.	□
The facilitator does not encourage participation or foster a sense of peer support.	□
The facilitator does not provide added value to the PLC.	□

29. How is the majority of the content of your PLC meetings determined?

Jointly by the PLC members	□
By the PLC leader or facilitator	□
By a scripted guide-manual	□

30a. Did you receive a guide or manual for the PLC meetings?

Yes	□
No	□

30b. If you received a guide or manual for the PLC meetings, which statement(s) are true about the materials provided? *(mark all that apply)*

The guide-manual provided is relevant and helpful.	□
The guide-manual provided helps guide the group discussion and/or activity during the PLC meeting.	□
The guide-manual provided helps me to reflect on my own teaching practice.	□
The guide-manual provided gives me new ideas for improving my teaching practice.	□
The guide-manual provided is *not* relevant or useful	□

31a. Do your PLC meetings include the participation of an external expert (for example, an expert in a subject area that is being taught, etc.)?

Yes	□
No	□

31b. If an external expert participates in your PLC meetings, which statement(s) are true about the expert? *(mark all that apply)*

The expert helps us to understand how to improve our teaching in a specific area.	□
The expert provides feedback on challenges posed by the group.	□
The expert enriches the discussion by providing specific examples from the expert's experience or from research.	□
The expert does not provide extra value to the group discussions.	□

32. Please indicate to what extent you agree or disagree with the following statements about the PLC meetings in which you participate

	Strongly agree	Agree	Not sure	Disagree	Strongly Disagree
My colleagues in the PLC share my values about children and their ability to learn.	□	□	□	□	□
My colleagues in the PLC share my values around the roles of teachers.	□	□	□	□	□
My colleagues in the PLC agree on shared norms about the use of time and space within the school.	□	□	□	□	□
The teachers in my PLC have similar teaching philosophies	□	□	□	□	□
The discussions in my PLC go beyond superficial exchanges and incorporate a deep discussion of practice.	□	□	□	□	□
In my PLC we collaboratively review student work student data as a way to improve instructional practices.	□	□	□	□	□
In my PLC, we focus the discussion on what and how well students are learning rather than how we are teaching.	□	□	□	□	□
In my PLC we focus on not simply ensuring that students are taught but on ensuring that they learn.	□	□	□	□	□

33. In a typical PLC meeting, about how much time is spent on the following…

	more than 85% of the entire meeting time	50 - 85% of the entire meeting time	25 - 49% of the entire meeting time	less than 25% of the entire meeting time
diagnosing individual students? Teachers discuss problems of specific students and arrange appropriate help.	☐	☐	☐	☐
analyzing teaching practices? Teachers discuss specific teaching practices of team members.	☐	☐	☐	☐
coordinating content? Teachers decide common themes, suggest related materials and activities to guide instruction	☐	☐	☐	☐
preparing lesson plans?	☐	☐	☐	☐
discussions about how to teach a particular concept?	☐	☐	☐	☐
preparing instructional materials?	☐	☐	☐	☐

34. How often since you have participated in PLC meetings starting in January 2018 have you:

	Never	Once a year or less	2-4 times a year	5-10 times a year	1-3 times a month	Once a week or more
Received meaningful feedback on your performance from your PLC colleagues?	◯	◯	◯	◯	◯	◯
Visited other PLC colleagues' classrooms to observe and discuss their teaching?	◯	◯	◯	◯	◯	◯
Had your PLC colleagues observe your teaching?	◯	◯	◯	◯	◯	◯
Received useful suggestions for curriculum materials from your PLC colleagues?	◯	◯	◯	◯	◯	◯
Invited a PLC colleague to help teach your class(es)?	◯	◯	◯	◯	◯	◯

35. Please indicate to what extent you agree or disagree with the following statements about the supportive conditions for your PLC meetings:

	Strongly agree	Agree	Not sure	Disagree	Strongly Disagree
There is time built into the school calendar for my PLC to meet.	☐	☐	☐	☐	☐
There is a physical space devoted for the PLC to meet without interruptions.	☐	☐	☐	☐	☐
There is a sufficient amount of time in each PLC meeting to accomplish goals.	☐	☐	☐	☐	☐
There is a sufficient amount of time to meet within my PLC to accomplish goals.	☐	☐	☐	☐	☐
The school director supports our PLC meetings.	☐	☐	☐	☐	☐
The PLC meeting place is close to my school.	☐	☐	☐	☐	☐
It is easy for me to travel back and forth the PLC meeting place.	☐	☐	☐	☐	☐

36. On a scale of 1 to 5 (with 1 meaning very negative and 5 meaning very positive), what rating best describes your feelings about the PLC meetings regarding each of the following aspects?

	1	2	3	4	5
Productive	□	□	□	□	□
Task oriented	□	□	□	□	□
Well facilitated	□	□	□	□	□
Collaborative	□	□	□	□	□
Compatible group members	□	□	□	□	□
Honest communication	□	□	□	□	□

37. To what extent has your participation in the PLC meetings directly led you to acquire any of the following?

	No positive change	A small change	A moderate change	A large change
New knowledge about instructional methods	□	□	□	□
New insights about how to meet individual needs of students	□	□	□	□
New ideas about how to improve the way you teach	□	□	□	□
New perspectives on your strengths and weaknesses in teaching	□	□	□	□
Greater confidence in using a wider range of instructional methods	□	□	□	□
A stronger sense of connection or support from other teachers	□	□	□	□
New ideas about how to enhance lesson plans	□	□	□	□
New ideas about how to manage classrooms effectively	□	□	□	□
New insights about how to use students' work to plan for instruction.	□	□	□	□
New insights about how to prepare instructional materials	□	□	□	□
New insights on how to deal with a challenge in my classroom	□	□	□	□

Chapter 8. Conclusion

8.1. Introduction

In order to improve teaching quality in Sub-Saharan Africa for in-service teachers and improve student learning, the literature recognizes that it is critical for support to teachers to be provided close to the school or within the school and be related to improving instruction (Bashir, 2018; Hardman et al., 2011; Hardman, 2009). The argument is that learning that occurs in the school as teachers and administrators engage in their daily work activities is likely to be most effective as it allows for concrete problems faced in the local environment to be raised, and teachers can reflect together and receive feedback on their actual teaching (Woods & McQuarrie, 1999). Professional Learning Communities (PLCs) allow for this type of learning by involving a group of teachers sharing and critically interrogating their practice in an ongoing, reflective, collaborative, inclusive, learning-oriented, growth-promoting way (Stoll and Louis, 2007; Stoll et al., 2006; Toole & Louis, 2002; Mitchell & Sackney, 2000). PLCs emphasize collaboration and reflection among a group of teachers to expose them to new ideas and practices, so they can improve on their pedagogy and their teaching through a process of critical inquiry. However, despite the recognized importance of within school continuing in-service professional development (including PLCs) still little is known on how collaborative efforts between teachers themselves and between teachers and school heads or coaches-mentors can influence teaching practices and student achievement.

In this dissertation, we contribute to the existing literature by investigating the influence or relationship between different types of within and across school collaboration practices on teaching and student learning. Specifically, we focus on (i) collaboration within school between teachers on pedagogical topics; (ii) principals' instructional support to teachers; and (iii) teachers' collaboration through PLCs.

In particular, this thesis included the following: (i) an analysis of the accuracy, reliability and validity of the data collected by the Equatorial Guinea Primary Education Survey (EG-PES) based on quality standards established in the survey methodology and student learning assessment literature; (ii) an empirical analysis of the effect of teachers' level of collaboration with their peers around pedagogical topics on student achievement; (iii) an analysis of the influence of principals' instructional support to teachers on student achievement; (iv) an exploration of how PLCs function in developing countries and how that differs from existing Western conceptualizations; (v) an exploration of the association between high levels of participation in PLCs and constructivist student-centered teaching; and (vi) an analysis of the relationship between the strength of a PLC and constructivist student-centered teaching.

From a methodological perspective, this thesis employed a mix of quantitative and qualitative methods. Quantitative methods included psychometric analysis as well as different econometric techniques to examine the relationships of interest. These included: instrumental variable estimations, pseudo-panel approach, coarsened exact matching, cluster robust inference, two-level intercept estimates and regression analysis with fixed-effects. The use of multiple methodologies and econometric techniques provide different levels of causal interpretation to similar research questions and allow us to test the evidence through different mechanisms. Qualitative methods included a review of the survey methodology and student learning assessment literature. Furthermore, this thesis also employed document analysis, semi-structured interviews and validation sessions with experts to dig deeper into how PLCs are working in developing countries – with a focus on Sub-Saharan Africa countries.

The remainder of this chapter is organized as follows. Section 8.2 articulates the main research findings through statements, which are substantiated by the evidence gathered in the dissertation. Section 8.3. covers study strengths and limitations and areas for future research. Finally, section 8.4. discusses the implication of the findings for policy.

8.2. Main Findings

This section presents the mains findings from the empirical chapters of this dissertation. The findings are presented in the form of seven statements. The first six statements are directly linked to the research questions highlighted in the introduction, which are substantiated by the evidence gathered in the individual chapters. The seventh and last statement combines the findings presented across the empirical chapters to inform the main take-ways of this dissertation.

Statement 1: Administration of a nationally-representative learning assessment in primary education in a low-and middle-income country can respond to quality standards and provide reliable data on student learning in Mathematics and Communications.

Chapters 2 of this thesis investigated if EG-PES complied with good standards highlighted in the survey methodology literature. As the Monitoring and Evaluation Specialist for a large education program responsible for leading all technical and logistical aspects of the EG-PES, the PhD candidate encountered a few challenges in implementing the EG-PES that may peculiar to developing countries.

First, a complete and up to date list of all schools in the country that could serve as a sampling frame was not available. This is often a challenge in developing countries, as school census data may not up to date and in some cases may not be available at all. In the case of EG-PES, the 2011 list of schools from the Primary Education School Census was used as the sampling frame instead. Because the Census was

conducted three years before the survey, the final sampling frame may underrepresent newer, and likely smaller schools, established in fast growing areas, such as the capital city (Malabo) and the city of Bata. In addition, there was no information available on the schools' ethnolinguistic compositions to use in the stratification process, so we could not ensure representation of ethnic groups in the final survey sample.

Second, we encountered challenges in recruiting enumerators representing different ethnic groups and with the right capacity level. Although we aimed to match students and interviewers based on the student's native language to avoid interviewer effects, we did not receive enough applications to the interviewer position from minority ethnic groups. This could have generated some interviewer related error, as it possible that a child who spoke Fang felt more comfortable with an interviewer whose native language was also Fang than a child who spoke Bulbi or Kombe. Additionally, performance of most candidates for the interviewer position was below satisfactory levels in the initial reading and mathematics exams, which raised questions about their ability to apply the student learning assessment, especially to advanced grades in primary education. Even after this initial screening, not all candidates who participated in the enumerator training demonstrated having acquired the skills in standardized interviewing necessary to apply the instruments. Because performance was lower than expected, we were unable to hire the targeted number of enumerators and had to adjust field logistics to increase the number of days allocated to data collection.

Finally, given the authoritarian and repressive nature of the political regime, one could think that fear could have been the main motivating factor to participate in the study. However, discussion with local staff indicated that it was highly unlikely that participation in the survey was driven by fear and the program's good imagine and standing within the education community likely lead to the high participation rates. During the survey implementation there were cases where local governments and-or school directors asked interviewers for financial resources or support to the school (textbooks, infra-structure, etc.). However, they did not condition their participation in the survey to the receipt of these benefits.

Despite these challenges that may be typical to applying surveys in a low-and middle-income country, and as discussed in detail chapter 2, the EG-PES complied with good standards, identified and attempted to minimize potential survey errors where possible. Given that the EG-PES followed best practices in sampling design, it is also fair to assume that the final sample represents the underlying population and that the empirical analysis conducted under the empirical chapters can be generalized to the population of grades 1 and 3 students in Equatorial Guinea

Finally, Chapter 3 provided a discussion on the development and psychometric properties of the EG-PES learning assessment instruments. The review of the process for instrument development and examination of its psychometric properties allowed us to conclude that the instruments and measures used follow agreed

standards for quality, validity and reliability of instruments. The learning assessment specifically provided reliable data on student learning in Mathematics and Communications.

Statement 2: Teachers' high level of collaboration around pedagogical topics has a positive effect on student achievement in language, but not in mathematics.

Chapter 4 investigated the impact of teachers' collaboration on student achievement in Equatorial Guinea using exogenous variation in the number of teachers per school as an instrument for the endogenous variable teachers' collaboration. Employing a 2SLS estimation with a latent-variable model applied in the first-stage of the 2SLS procedure, the instrumented variation in teachers' collaboration is used to carve out the impact on student achievement. Our results showed that, on average, teachers that exhibit a high level of collaboration cause an increase of approximately 0.5 standard deviations in the score in the language assessment, which is regarded as a large effect by the education literature, even in developing countries. There are two possible reasons for this large effect size. First, one can expect that effect sizes are larger in developing countries as educational innovations are less prone to ceiling effects, which are typically observed in developed countries. Second, the large effect observed might be related to the binary nature of the variable collaboration, which measures two extremes of collaboration – high versus low levels of collaboration.

Our results do not show a significant effect of teachers' collaboration in the mathematics assessment. Unfortunately, we lack data on the nature and focus of collaboration between teachers, which may cover topics on mathematics instruction, reading instruction or other. Given the Ministry of Education's focus on improved reading since 2006 and the development and distribution of materials to improve reading instruction, it is likely that teachers' collaborative activities placed a higher focus on language. This would explain why we find a significant effect of teachers' collaboration in language, but not in mathematics.

Our results are stronger when we restrict our sample to Fang teachers only: we observe that the influence of number of teachers in primary grades on the probability of teachers exhibiting high levels of collaboration is higher among the restricted sample of Fang teachers. These results indicate that teachers' background may influence their likelihood to collaborate, as teachers may be more willing to collaborate with other teachers that share the same background characteristics.

Statement 3: High levels of participation in PLC is associated with constructivist student-centered practices.

Chapter 7 examines the relationship between participation in PLCs and constructivist student-centered teaching using data from a teacher survey implemented with over 2,300 preschool and primary education

teachers in Equatorial Guinea, Ghana and Nigeria. Results from Ordinary Least Square (OLS) regressions show a positive significant relationship between higher levels of participation in PLC and constructivist student-centered practices in Nigeria and Ghana, but not in Equatorial Guinea. Equatorial Guinea was the only country for which we were unable to include school-fixed effects in the regression analysis. Unobserved and observed school factors influencing the teaching outcome of interest that we were unable to control for could have affected our results. On average, teachers who exhibit high participation in PLCs are associated with an increase of approximately 0.392 standard deviations in the constructivist student-centered teaching score in Ghana and 0.316 standard deviations in Nigeria. Results from a sensitivity analysis suggest that one or two meetings a month may not be sufficient to drive associations with constructivist student-centered teaching and that higher levels of participation in PLCs may be needed to have an influence on the teaching outcomes of interest.

Our results may suggest that teachers' interaction in PLCs was working to change existing traditional teaching practices towards more active pedagogical approaches. This is an important finding in light of the debate in the literature on whether PLCs promote reformation or transformation of practice (Lee & Lee., 2018; Philpott & Oates, 2017; Servage, 2008). Reformation of practice involves aligning individual teaching practices with prescribed norms or existing mandated goals (Philpott & Oates, 2017), whereas transformation of practice involves questioning goals and empowering teachers to be *"... producers of pedagogical knowledge rather than solely consumers and implementers*" (Philpott & Oates, 2017, p. 213). The association we find may indicate a transformation of practice towards more student-centered approaches in contexts characterized by high teacher-centered instruction.

Statement 4: Instructionally supportive principals have a positive influence on student achievement in mathematics, but not in language.

Chapter 5 investigates the influence of instructionally supportive principals on student achievement, taking advantage of a process that mimics a natural experiment in Equatorial Guinea, according to which nearly all principals are hired by the Ministry of Education and Science and placed arbitrarily in schools across the country. Using a Coarsened Exact Matching approach to construct a pseudo-panel, third graders are matched with first grade students based on key observable characteristics, which allows us to control for estimated prior achievement. Our results showed that there is a positive and significant influence of instructionally supportive principals on the performance of students in mathematics. On average, principals that provide high support to teachers on instructional topics are associated with an increase of approximately 0.2 standard deviations in the score in the mathematics assessment. This is regarded as a medium effect by the education literature and is in line with earlier research, highlighting that teacher effects tend to be higher than principal effects in magnitude. Our results did not show a significant influence of instructionally

supportive principals on the performance of students in language. This finding is also consistent with previous literature that found a positive relationship between principal experience and math test scores, but not between principal experience and English scores (Clark et al., 2009) or found a significant influence of principals in both subject areas, but the magnitude is higher for mathematics than for reading (Dhuey & Smith, 2014; Alig-Mielcarek & Hoy, 2005). A heterogeneity analysis by urban and rural areas show a positive and significant influence of instructionally supportive principals in urban areas on students' score in the mathematics assessment, but no significant influence is found for principals is rural areas. This result is not surprising, given acute conditions of schools in rural areas, which may make it harder for principals to have an influence on student learning through instructional support to teachers.

Statement 5: The mode of PLCs operation in Sub-Saharan African countries differs from Western conceptualizations and reflects structured and scripted PLC models.

In a recognition that existing conceptualizations of PLCs from the Western literature may not reflect how PLCs are functioning in developing countries, Chapter 6 aimed to inductively create a typology of PLCs that incorporates elements that might be specific to these countries, with a focus on Sub-Saharan Africa. A multimethod approach, encompassing document analysis, semi-structured interviews with PLC experts and expert validation, allowed us to systematically review and evaluate documents on PLCs in Sub-Saharan African countries and triangulate the findings through semi-structured interviews with experts and practitioners working on PLCs. Three models of PLC operation in Sub-Saharan African countries emerged from our research, – *autonomous, structured and scripted* – forming a PLC typology. The typology reflects how PLCs function in low- and middle-income countries, and in Sub-Saharan African countries in particular, but also holds for developed countries. While in low- and middle-income countries the *structured* and *scripted* models of PLC prevail, in developed countries the *autonomous* model is more commonly found. Our document analysis revealed that within any of the three models specific features of PLCs can vary greatly, so we also account for *structural varying features*, which refer to all the distinctive features associated with how PLCs are designed that might vary across PLCs. There are also key differences between high and low-resource contexts in terms of *structural varying features*. For instance, while PLCs are described in the literature almost exclusively as school-based, in developing countries they are often comprised by a group of schools previously organized through the education system.

We integrate our typology of PLCs and corresponding structural varying features with key elements proposed by the Western literature, forming one cohesive conceptual framework. The framework is organized around five interrelated PLC dimensions, three of which emerge from the Western literature (*core characteristics*, *supportive conditions* and *phases of development)* and two (PLC *models and*

structural varying features) which emerged from our document review of PLCs in low- and middle-income countries and semi-structured interviews with experts, validated through expert discussion sessions.

Statement 6: Stronger PLCs are associated with higher constructivist student-centered teaching.

Chapter 7 examines the relationship between the strength of a PLC and constructivist and student-centered teaching practices in Nigeria, Ghana and Equatorial Guinea. To explore this association, we first built a PLC Strength Index based on the PLC core characteristics from the conceptual framework we propose in Chapter 6. An exploratory factor analysis revealed that four out of the five core characteristics measure one overarching construct of PLC and these four characteristics are used to construct the Index. Results from OLS regressions with cluster-robust standard errors demonstrated that across all countries under study, stronger PLCs are significantly associated with higher constructivist student-centered teaching. The regression analysis showed that each point in the PLC Strength Index (which means adding one core PLC characteristic) is associated with an increase of 0.118 standard deviations in constructivist student-centered teaching in Ghana, 0.112 in Nigeria and 0.192 in Equatorial Guinea. This finding indicates that these four core characteristics conceptualized by the Western literature may be applicable in the context low-and middle-income countries.

Chapter 7 also explored specific mechanisms that interact with PLC strength to influence teaching. We expanded the original regression model to include three structural varying characteristics of PLCs from typology presented in Chapter 6 – presence of an active facilitator, content of the meetings determined by a guided-manual, content of the meetings determined jointly by PLC members - and explored how these features interact with PLC strength to influence constructivist student-centered teaching. Descriptive statistics showed that the lower the PLC strength, the higher the benefits of having an active facilitator and a guided-manual (as per the scripted model) to determine the content of a PLC meeting on constructivist student-centered practices. Finally, in the specific case of Nigeria, having the content of the meeting be determined jointly by PLC members in weak or medium strength PLCs is negatively associated with our teaching outcomes of interest.

Statement 7: Within schools and across schools' collaborative practices and structures in Equatorial Guinea, Ghana and Nigeria that have the potential to improve teaching and learning.

The findings from the four empirical chapters combined indicate that collaboration within and across schools among teachers as well as between principals and teachers can influence teaching practices and student achievement. Teachers who collaborate more often in pedagogical matters positively effect student language achievement while teachers who participate more often in within school and across schools PLCs are more likely to exhibit constructivist student-centered practices. Additionally, we found that principals

that provide high support to teachers on instructional topics are associated with an increase in student mathematics achievement. These results suggest that teacher's collaboration within and outside PLCs as well as collaboration with the principal may work to change traditional teaching practices towards more active pedagogical approaches that can influence student achievement. Further research is needed to explore in more depth the differences observed by subject matter (principals' support influence mathematics achievement while teachers' collaboration affect language achievement). This research also explored if PLCs function in a different way in low-and-middle income countries. Our typology of PLCs captured structured and scripted models of PLCs that may be more common in these countries, while the autonomous model seems to prevail in high-income countries. However, we found that at least four of the five characteristics for a successful PLC highlighted in the Western literature may also be valid in low-resources countries.

8.3. Strengths and Limitations

Although each of the four empirical chapters of this dissertation have already highlighted specific strengths and limitations, there are still some general strengths and limitations of this research that we highlight in this section.

Strengths

This dissertation focused on studying Equatorial Guinea, a country for which there is no published literature critically investigating the education sector. Given the restricted political rights and civil liberties in the country, overall research has not been encouraged and often not allowed by the government. This dissertation also moves beyond the case study of one single country to include Ghana and Nigeria, which allows for a comparative perspective within Sub-Saharan Africa. As a research and evaluation specialist working for a major non-profit organization, I was able to promote and lead primary data collection efforts to provide firsthand insight into learning and teaching outcomes in these three countries. The unique data collected, albeit not perfect, has allowed for in-depth studies of teachers' collaborative practices, principal's instructional support and functioning of Professional Learning Communities, which are often under-studied, especially in the context of low-and middle-income countries.

Limitations

First, data used on the analysis is based on frequency of collaborative practices – frequency of teachers' collaboration on pedagogical topics, frequency of principal's instructional support to teachers, frequency of teachers' participation in PLCs - and does not contain information on the nature of the collaborative processes. The distinction between task effectiveness and behavioral frequency is an important one

(Grissom & Loeb, 2011). As Hallinger and Murphy (1985) explain, ''*certain behaviors could be performed frequently but in a perfunctory or ritualistic manner [while] certain practices probably do not need to be performed frequently in order to be performed effectively*'' (p. 226). Collaboration among teachers and between teachers and principals can vary greatly, ranging from superficial exchanges to in-depth reflection and discussion of teaching practices and analysis of student work. Without this information, we cannot know which specific interactions and dynamics of collaboration may lead to teacher learning, innovations in teaching practices and improved student learning. Further research investigating the channels through which teacher collaboration and principal instructional support can lead to student learning is warranted.

Second, we rely on teachers' self-reported data to construct a number of variables of interest, including the use of constructivist student-centered practices in the classroom and the PLC Strength Index. Even though the survey data was kept confidential and it was clearly explained to teachers that no incentive would result from the survey, unobserved characteristics, such as ambition or ability, could be related with the likelihood of over reporting. Future research that relies on observational data of teaching practices and of PLC strength is recommended to fill out this gap.

Third, some of the methodologies used in this dissertation do not allow for a causal interpretation. For example, the analysis between high levels of participation in PLC and constructivist student-centered practices as well as between PLC strength and these teaching outcomes only provide correlational results. Our goal with this analysis is to document empirically the linkages between our variables of interest, and we cannot say, for example, that PLCs causes constructivist student-centered practices. However, given the lack of quantitative research on PLCs, even an exploratory study can make a valuable contribution to the literature. While we cannot prove a causal hypothesis with the type of correlational analysis used here, we can rule one out. Overall, more longitudinal as well as experimental or quasi-experimental research on PLCs is warranted.

Finally, despite our systematic search for documents on PLCs, the PLC typology that emerged from the qualitative research may not include the universe of existing documents on PLCs in Sub-Saharan Africa. Additionally, because we adopted a convenience sample selection approach to the semi-structured interviews and validation sessions, we may be missing potentially different views and perspectives of experts working on PLCs. Further research testing the validity of the new empirically based typology and integrated conceptual framework is warranted. In addition, in order to expand the validity of our typology into other settings, more research is needed exploring other regions of the world and other educational levels (e.g., secondary and tertiary).

8.4. Policy Implications

Over the past two decades the international debate on education has shifted from prioritizing educational enrollment only to also including educational quality. While the 1990 World Education For All (EFA) conference in Thailand focused on achievement of universal primary education by the end of the decade, the Dakar EFA conference in 2000 made reference not just to enrollment, but to the quality of education pupils would receive (Courtney, 2008). All three sub-targets elaborated in Dakar, while focused on universal education, made explicit reference to quality. In 2005, the EFA Global Monitoring Report was published with a targeted focus on quality. The report acknowledged that the achievement of universal participation in education will be fundamentally dependent upon the quality of education available (UNESCO, 2005).

In 2015, at the United Nations Sustainable Development Summit in New York, Member States formally adopted the 2030 Agenda for Sustainable Development. The agenda, which contains 17 goals, included a new goal on global education (SDG 4) with an explicit focus on quality: *"ensure inclusive and equitable quality education and promote lifelong learning opportunities for all"* (UNESCO, 2016, p.4). One of the seven targets under SDG4 reads "*By 2030, ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes*" (UNESCO, 2016, p.33). This target includes sub-targets that focus exclusively on learning outcomes: on achieving at least a minimum proficiency level in reading and mathematics and in measuring these through administration of nationally representative learning assessments. One of the three means of implementation proposed under SDG 4 recognizes that teachers are the key to achieving all of the SDG4 targets and calls for the provision of pre-service or in-service training to teachers: "*By 2030, substantially increase the supply of qualified teachers, including through international cooperation for teacher training in developing countries, especially least developed countries and small island developing States*" (UNESCO, 2016, p.15).

The current debate recognizes that many low- and middle-income countries have made enormous progress in the past decades in expanding enrollment and in achieving the target of universal primary education (World Bank, 2018; UIS, 2016). Nonetheless, the expansion in schooling has often been accompanied by a learning crisis in these countries, characterized by a higher proportion of children attending school, but failing to learn the expected competences for their grades (UNESCO, 2013). Indeed, a large share of children in low-income countries complete their primary education lacking basic reading, writing, and arithmetic skills (Bold et al., 2017).

In light of the international focus of going beyond access to school to improving education quality once students are in school, the findings of this dissertation provide some indicative policy implications on how to improve the quality of education through sustainable improvements in quality of teaching and in the quality of school principals. In Equatorial Guinea, collaboration among teachers on pedagogical matters improve student achievement. This collaboration is stronger among teachers from the same ethnic group and yields stronger impact on student achievement, In Ghana and Nigeria, teachers who participate frequently on PLCs are more likely to exhibit constructivist student-centered teaching. In Equatorial Guinea, highly instructionally supportive principals influence student learning. Taken together, these findings support policies and regulations that focus on the establishment and growth of within schools and across schools' collaborative structures as a form of continuous in-service professional development in Equatorial Guinea, Ghana and Nigeria that have the potential to improve teaching and learning.

It is not possible to know the external validity of these findings to other countries in Sub-Saharan Africa. Nonetheless, similarities in trends as well as challenges faced by the different education systems across the continent make these findings also relevant to other countries in the region. Across Sub-Saharan African countries, students are failing to reach the minimum level of learning, as defined by the various regional and international assessments (Bashir et al., 2018). Teachers in the region often have low educational qualifications (UIS, 2006), do not master the curricula, have low basic pedagogical knowledge and rarely use good teaching practices (Bold et al., 2017). Most often in-service teacher training is delivered through short courses centrally planned by the Ministry of Education on specific topics, such as the introduction of new curricula (Lauwerier & Akkari, 2015), and are not connected to the context or the on-the-ground reality teachers experience in their school communities.

Despite these challenges, some common trends are also starting to emerge towards more flexible school-based provision of professional development in the region (Hardman et al., 2011, Hardman et al., 2009, Mattson, 2006). Several Ministries of Education have established internal departments and units which work through their decentralized local offices at the regional, district and zonal-level to monitor and support school-based programs (Mattson, 2006) and have set-up national in-service education and training (INSET) strategies and continuous professional development (CPD) systems for teachers (Hardman et al., 2011). Overall, some countries are starting to move away from ad-hoc provision to a more systematic and longer-term approach where the teacher is more involved in his or her on-going professional development, and where other actors play critical supporting roles (Hardman et al., 2011).

By showing that collaboration among teachers in the same school or cluster of schools and between teachers and principals can improve teaching and learning, the findings from this research support the trend observed in some countries in the region towards a long-term sustainable vision of continuous professional

development based on flexible approaches grounded in the school or cluster of schools. The results indicate that a shift away from centralized control to decentralized delivery of in-service professional development that is based on collaboration between teachers and the school leadership on pedagogical topics may be possible.

Policy recommendations emanating from this dissertation include the need for countries in Sub-Saharan Africa to diversify approaches to in-service education to go beyond on-off trainings and include more school-based approaches to teacher professional development, which this dissertation and other emerging literature (Conn, 2014) are showing to be effective and relevant to teachers' classroom needs (Junaid & Maka, 2015). This includes policies and programs at national or district level to support the establishment of school-based or across schools PLCs and the provision of capacity building opportunities to principals (and other school leader) so they can effectively coach and mentor teachers on instructional matters. We discuss both approaches in more detail below. However, although our findings support these policies, more research is needed to understand what this means in practice in terms of organizational demands of decentralized delivery, administration and support (Mattson, 2006), on how to stablish elaborate structures for effective management and governance of a decentralized INSET system (Junaid & Maka, 2015) and on how to ensure teachers have the time, support and resources needed to participate in decentralized in-service education. Finally, we recognize that these policies and programs are just one aspect of the puzzle and cannot alone fully improve education quality. Policies affecting other contextual conditions that may influence learning (student's nutrition, school infra-structure, availability of materials, school environment, pupil-to-teacher ratio, etc) should also be considered for maximum impact on education quality.

Establishment of PLCs

Findings provide strong support for national or district level policies and programs that establish PLCs as a form of school-based and cluster-based ongoing professional development. Although there are still many questions to be answered around the optimal design and implementation of PLCs in different contexts, the new empirically based typology and integrated conceptual framework presented in this dissertation can offer initial guidance to policy-makers and practitioners on the design and promotion of these collaborative structures. In attempts to answer some of the question on PLC design, we tested empirically specific elements of the conceptual framework introduced in Chapter 6. Our results underline the importance of policies and programs that aim to promote strong PLCs and create a common vision focused on fostering at least four out of five core characteristics of PLCs from our framework in conjunction, not separately.

We also explored mechanisms through which PLC strength can influence teaching. Our results suggest that a scripted model may be helpful in supporting the work of PLCs that are just starting or are weak

through the development of materials that can help stimulate and guide teachers' work. More importantly, these results may indicate that a scripted model of PLCs can lead to transformative teaching practices as opposed to alignment of teaching with prescribed norms or mandated goals. On the contrary, policies and programs that provide full autonomy to PLCs that are starting may not be desirable, given the negative association found between content determined jointly by PLC members and the teaching outcomes of interest for PLCs that are still weak. A direct policy recommendation that emanate from these findings include policies or programs that promote scripted models of PLCs where PLCs are a new practice among the teaching community and are just starting. This includes the development of context-specific scripted manuals that will guide participating teachers, encourage and stimulate the discussion, at the same time as promoting the four core PLC characteristics found key for promoting a change in teaching practices.

New PLC policies and programs should also train facilitators on active facilitation techniques as PLCs start for influencing teaching practices. Although there is a need for more research focusing attention on how policies and programs can promote and sustain the four core characteristics, the results from the exploration of mechanisms may indicate that a scripted model and a trained facilitator may be helpful resources in this process. However, as PLCs become strong on the four core characteristics, programs and policies may need to move away from scripted approaches to more fully autonomous model in order to have an effect on teaching. Hence, there is no one golden policy recommendation. Instead, we suggest that policies and programs acknowledge that PLCs are not static and evolve over time, and that fluid programs and policies are needed to respond to changing PLC needs.

The findings from this research, which indicated that in Equatorial Guinea collaboration was stronger among teachers from the same ethnic affiliation, may support the establishment of PLCs among specific teacher sub-groups. While designing PLCs to promote collaboration among teachers from the same ethnicity or religious affiliation may be a discriminatory practice, some attention to other grounds of solidarity may be given during PLC design. Previous literature has already recognized that professional communities are more likely to occur in homogeneous rather than heterogeneous groups (Bryk et. al., 1999). Although more research is needed on PLC composition, PLCs may consider focusing on specific grades, subject areas or themes of interest, as a way to promote these homogeneities through shared interests. Similarly, PLCs across schools may consider clustering schools that share the same instructional-pedagogical beliefs and practices, or schools that share similar challenges.

Finally, decision-makers should consider context relevant supportive conditions to PLCs. Although we do not explore the role of supportive conditions empirically as part of this research, the Western literature highlights that supportive conditions are key to inhibiting or sustaining PLCs. Supportive conditions that decision-makers may consider under PLC policies and programs include: built-in meeting time within the

school calendar for teachers to meet; provision of incentives (such as professional development credits); and training of principals as school leaders that fully support PLCs within their schools.

Development of principal's instructional leadership capacity

Finally, our findings support policies that invest in professional development activities that aim to foster principal's instructional leadership. Previous studies have found that principals who participate in instructional leadership trainings are more engaged in instructional leadership (OECD, 2016). Instructional leadership can be fostered by encouraging principals to be up to date with developments in their field through in-service training and through attendance of professional development activities. School-based mentoring and support to principals should be considered as a central mechanism of INSET strategies as a way of building technical capacity of school leadership. Instructional leadership trainings to principals may also have the potential to decrease the urban-rural gap in student achievement. To do so, instructional leadership trainings should be tailored to the context of rural schools and equip principals with instructional support practices that aim to tackle challenges that are specific to this context.

8.5. Directions for Future Research

This thesis provides direction for further research. First, experimental research in both high and low resources contexts is warranted to rigorously evaluate the impact of PLCs on teaching practices and student outcomes. An experimentally design study could overcome one of the main challenges faced by this dissertation, which is attributing causality. We recommend combining any experimental design with a qualitative approach to disentangle the mechanisms through which PLCs influence teaching. This could involve both direct observations of PLC internal dynamic as well semi-structured interviews with participant teachers.

Since PLCs are imbedded in a social, political, cultural and economic context, the typology proposed in this thesis opens the door for future research on which PLC model is best for different settings. For instance, in contexts characterized by high power distance, hierarchy and authoritarian education systems, which PLC model can more effectively facilitate changes in instructional practice? This question could be approached through an evaluation design with different treatment arms, randomizing different PLC models among an eligible population of schools (or cluster of schools). Additionally, through experimental designs further research should investigate which specific PLC structural varying features are most effective in promoting teacher effectiveness.

Finally, given our finding related to the influence of principals' instructional support to teachers on student performance, research investigation the impact of training programs to principals focusing on strengthening

their instructional leadership capacity is warranted. There is only a small body of evidence investigating the impact of these interventions in developed countries, and no rigorous evaluation has been conducted in low- and middle-income countries. We need to further understand if principal instructional leadership capacity can be promoted through training programs and whether this strengthened capacity is translated into improved teaching and student outcomes.

References

Alig-Mielcarek, J., & Hoy, W. K. (2005). *Instructional leadership*. Information Age Publishers: Greenwich, CT, USA.

Bashir, S., Lockheed, M., Ninan, E., & Tan, J. P. (2018). *Facing forward: Schooling for learning in Africa*. Washington, DC: World Bank.

Bold, T., Filmer, D., Martin, G., Molina, E., Stacy, B., Rockmore, C., ... & Wane, W. (2017). Enrollment without learning: Teacher effort, knowledge, and skill in primary schools in Africa. *Journal of Economic Perspectives*, *31*(4), 185-204.

Clark, D., Martorell, P., & Rockoff, J. (2009). School Principals and School Performance. Working Paper 38. *National Center for Analysis of longitudinal data in Education research*. Conn, K. M. (2014). *Identifying Effective Education Interventions in Sub-Saharan Africa: A meta-analysis of rigorous impact evaluations* (Doctoral dissertation, Teachers College).

Conn, K. M. (2014). *Identifying Effective Education Interventions in Sub-Saharan Africa: A meta-analysis of rigorous impact evaluations* (Doctoral dissertation, Teachers College).

Courtney, J. (2008). Do monitoring and evaluation tools, designed to measure the improvement in the quality of primary education, constrain or enhance educational development?. *International Journal of Educational Development*, *28*(5), 546-559.

Dhuey, E., & Smith, J. (2014). How important are school principals in the production of student achievement?. *Canadian Journal of Economics/Revue canadienne d'économique*, *47*(2), 634-663.

Grissom, J. A., & Loeb, S. (2011). Triangulating principal effectiveness: How perspectives of parents, teachers, and assistant principals identify the central importance of managerial skills. *American Educational Research Journal*, *48*(5), 1091-1123.

Hallinger, P., & Murphy, J. (1985). Assessing the instructional management behavior of principals. *Elementary School Journal, 86*(2), 217-247.

Hardman, F., Abd- Kadir, J., Agg, C., Migwi, J., Ndambuku, J., & Smith, F. (2009). Changing pedagogical practice in Kenyan primary schools: the impact of school- based training. *Comparative Education*, *45*(1), 65-86.

Hardman, F., Ackers, J., Abrishamian, N., & O'Sullivan, M. (2011). Developing a systemic approach to teacher education in sub-Saharan Africa: Emerging lessons from Kenya, Tanzania and

Uganda. *Compare: A Journal of Comparative and International Education*, *41*(5), 669-683.

Junaid, M. I., & Maka, F. (2015). In-service teacher education in Sub-Saharan Africa: A synthesis report. *Commonwealth Secretariat. London. Ministry of Education (1996), Educating Our Future: National Policy on Education, Lusaka: Zambia*.

Lauwerier, T., & Akkari, A. (2015). Teachers and the quality of basic education in sub-Saharan Africa. *Education Research and Foresight: Working Papers Series*, (11).

Lee, D. H. L., & Lee, W. O. (2018). Transformational change in instruction with professional learning communities? The influence of teacher cultural disposition in high power distance.

Mattson, E. (2006). *Field-based models of primary teacher training: case studies of student support systems from Sub-Saharan Africa*. London: DFID.

OECD (2016), *School Leadership for Learning: Insights from TALIS 2013*, TALIS, OECD Publishing, Paris, https://doi.org/10.1787/9789264258341-en.

Philpott, C., & Oates, C. (2017). Professional learning communities as drivers of educational change: The case of learning rounds. *Journal of educational change*, *18*(2), 209-234.

Servage, L. (2008). Critical and transformative practices in professional learning communities. *Teacher education quarterly*, *35*(1), 63-77.

UNESCO Institute for Statistics (UIS). (2006). *Teachers and Educational Quality: Monitoring Global Needs for 2015*. Montreal, Canada: Author.

UNESCO Institute for Statistics (UIS). (2016). *Leaving No One Behind: How Far on the Way to Universal Primary and Secondary Education?*. (Policy Paper 27/Fact Sheet 37). Montreal, Canada: Author.

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2013). *The Global Learning Crisis: Why Every Child Deserves a Quality Education*. Paris, France: Author.

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2016). *Unpacking Sustainable Development Goal 4–Education 2030*. Montreal, Canada: Author.

Woods, F., & McQuarrie, F. (1999). On the job learning. *Journal of Staff Development, 20*(3), 10-13.

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

Addendum on valorisation to the dissertation

The room is crowed. 52 third-grade students and one teachers to be exact. It is hot and humid, thirty-six degrees Celsius inside of the room but it feels like thirty-eight. The teacher asks one student, Joaquin, to come to the front of the classroom and read a simple sentence from the blackboard. Joaquin, at age 10, is not much older than his classmates. Joaquin entered school late, starting first grade at 8 years-old. Joaquin wakes up early every day to walk approximately five kilometers to get to his school. His parents work hard and save all they can to pay for Joaquin's school fees, materials and uniform. As Joaquin stands in front of the classroom, he is unable to read beyond the first word of the sentence. Feeling embarrassed, he gets back to his seat.

Joaquin attends an urban public school in Equatorial Guinea. Urban schools in Equatorial Guinea are often crowed and have high student-teacher ratios. Teachers have low qualifications, frequently employ traditional teaching methods and receive limited professional development opportunities. Joaquin is in third grade but lacks basic reading skills and is unable to read a single word of a short text. The traditional teaching practices employed by his teacher often makes him embarrassed, lower his self-esteem and makes him more prone to dropping out of school.

Over the past two decades the international debate on education has shifted from prioritizing educational enrollment only to also including educational quality. The current debate recognizes that many low- and middle-income countries have made enormous progress in the past decades in expanding enrollment and in achieving the target of universal primary education (World Bank, 2018; UIS, 2016). Nonetheless, the expansion in schooling has often been accompanied by a learning crisis in these countries, characterized by a higher proportion of children attending school, but failing to learn the expected competences for their grades (UNESCO, 2013). Indeed, a large share of children in low-income countries complete their primary education lacking basic reading, writing, and arithmetic skills (Bold et al., 2017).

The 2030 Agenda for Sustainable Development adopted in 2015 by United Nations Member States contains 17 goals, including a new goal on global education (SDG 4) with an explicit focus on education quality: ***"ensure inclusive and equitable quality education and promote lifelong learning opportunities for all"*** (UNESCO, 2016, p.4). One of the seven targets under SDG4 reads "*By 2030, ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes*" (UNESCO, 2016, p.33). This target includes sub-targets that focus exclusively on learning outcomes: on achieving at least a minimum proficiency level in reading and mathematics and in measuring these through administration of nationally representative learning assessments.

There is also an increasing recognition that the learning crisis is a teaching crises and to be effectively addressed we need to support teachers, who are the most important drivers of learning once students are in school. In fact, one of the three means of implementation proposed under SDG 4 recognizes that teachers are the key to achieving all of the SDG4 targets and calls for the provision of pre-service or in-service training to teachers: "*By 2030, substantially increase the supply of qualified teachers, including through international cooperation for teacher training in developing countries, especially least developed countries and small island developing States*" (UNESCO, 2016, p.15).

While many education systems still focus on ad-hoc and isolated provision of in-service teacher training, sustainable long-term approaches to teachers' professional development grounded in the schools are starting to emerge across low-and-middle income countries. In light of the international focus of going beyond access to school to improving education quality once students are in school, this dissertation examines how to improve the quality of education through sustainable improvements in quality of teaching and in the quality of school principals. Specifically, this dissertation investigates the influence or relationship between different types of within and across school collaboration practices on teaching practices and student learning. The innovative character of this dissertation lies in empirically exploring the effects of these approaches, which have been under-explored by the literature, through rigorous and innovative analytical methods. To the best of my knowledge, this is the only study assessing the influence of teachers' collaboration on student achievement. It is also the only study assessing the influence of principals on student achievement through the provision of instructional support to teachers in low-and middle-income countries.

In particular, this thesis analyzed: the effect of teachers' level of collaboration with their peers around pedagogical topics on student achievement; the influence of principals' instructional support to teachers on student achievement; the association between high levels of participation in Professional Learning Communities (PLCs) and constructivist student-centered teaching; and the relationship between the strength of a PLC and constructivist student-centered teaching. It also explored how PLCs function in developing countries and how that differs from existing Western conceptualizations. Overall, this research expands the understanding on how within school and across schools' collaboration practices can improve teaching and student learning. It provides empirical evidence showing that collaborative efforts between teachers themselves and between teachers and school heads can indeed improve teaching practices and student achievement.

The main research setting that is common across all empirical chapters in this thesis is Equatorial Guinea. The country's education system is characterized by low teachers' qualifications, traditional pedagogical approaches and low student learning outcomes (Bassett et al., 2017). Exploring alternative and sustainable

methods to teacher professional development that can lead to improved teaching and learning is key for the country to improve the delivery of quality education. Additionally, as Monitoring and Evaluation Specialist for the National Program for Improving Education Quality in Equatorial Guinea, I was able to design and lead data collection efforts for initial grades of elementary schools, which resulted in reliable and detailed data on student achievement in math and language as well as on student's, teacher's and principal's background characteristics. Although this dissertation focuses mostly on Equatorial Guinea, the challenges the country faces in primary education are similar to the challenges faced by other Sub-Saharan African countries (e.g., low student learning, low teachers' qualifications, teachers' use of traditional pedagogical approaches in the classroom, among others). The last empirical chapter of this dissertation also expands beyond Equatorial Guinea to collect data on teachers and their participation in PLCs in Ghana and Nigeria.

Three key findings from this dissertation include the following:

Finding 1: Teachers' high level of collaboration around pedagogical topics has a positive effect on student achievement in language

In Equatorial Guinea, collaboration among teachers on pedagogical matters improve student achievement in language. Teachers who collaborate more often in pedagogical matters positively affect student language achievement. This collaboration is stronger among teachers from the same ethnic group and yields stronger impact on student achievement. Finding 1 is particularly relevant in EG because of teachers' low level of qualifications and challenges in increasing teaching quality. By showing that collaboration among teachers in the same school can improve student learning, the findings from this research support the trend observed in some countries in the region towards a long-term sustainable vision of continuous professional development based on flexible approaches grounded in the school or cluster of schools.

Finding 2: Instructionally supportive principals have a positive influence on student achievement in mathematics

In Equatorial Guinea, highly instructionally supportive principals influence student learning indirectly, through their influence in teaching. This result supports government provision of capacity building opportunities to principals (and other school leader) so they can effectively coach and mentor teachers on instructional matters. Instructional leadership can be fostered by encouraging principals to be up to date with developments in their field through in-service training and through attendance of professional development activities. School-based mentoring and support to principals should also be considered as a way of building technical capacity of school leadership.

Finding 3: High levels of participation in PLC is associated with constructivist student-centered practices.

In Ghana and Nigeria, teachers who participate frequently on PLCs are more likely to exhibit constructivist student-centered teaching. Finding 3 informs us that more countries beyond Equatorial Guinea are likely to experience the benefits of teachers' collaboration and participation in PLCs. Therefore, we expect that most countries in Sub-Saharan country could benefit from the results and recommendations of this study. Finding 3 suggests that teacher's participation in PLCs may work to change traditional teaching practices towards more active pedagogical approaches that have the potential to influence student learning. This finding also indicate that a shift away from centralized control to decentralized delivery of in-service professional development that is based on collaboration between teachers may be possible.

Overall policy recommendations emanating from this dissertation include the need for countries in Sub-Saharan Africa to diversify approaches to in-service education to go beyond on-off trainings and include more school-based approaches to teacher professional development, which this dissertation and other emerging literature (Conn, 2014) are showing to be effective and relevant to teachers' classroom needs (Junaid & Maka, 2015). This includes policies and programs at national or district level to support the establishment of school-based or across schools PLCs. Although there are still many questions to be answered around the optimal design and implementation of PLCs in different contexts, the new empirically based typology and integrated conceptual framework presented in this dissertation can offer initial guidance to policy-makers and practitioners on the design and promotion of these collaborative structures.

The findings and recommendations from this PhD are, or at least should be, of interest to education government authorities at different levels (national, regional, local), international nonprofit organizations, multilateral organizations, and *specially,* to every single individual in education systems with a stake in learning, including teachers and school heads. It should be of interest to teachers and school heads as it may help them realize the power of collaborating with and supporting their peers. It should be of interest to international nonprofit organizations and multilateral organizations as they work with national governments in low-and-middle income countries to define systems and strategies for teachers' in-service education and continuous professional development that move away from ad-hoc provision of in-service teacher training to a more systematic, sustainable and longer-term approach grounded in the school or cluster of schools. While building a typology of PLCs this research both collaborated with and aimed at professional users from these development organizations, who may not have a scientific training, but derive their expertise and their commitment to this topic (PLCs) largely from professional experience. The new empirically based typology and integrated conceptual framework of PLCs presented in this dissertation can offer initial guidance to policy-makers and practitioners on the design and promotion of these collaborative structures. Finally, this research should be relevant to education government authorities as it provides supportive evidence to decentralize educational administration, management and oversight of in-service professional

development, provision of resources and delegation of responsibilities to local offices at the regional, district and zonal-level to monitor and support school-based programs.

Beyond societal relevance, this study also clearly marks an innovation in the in-service teacher professional development literature and teacher training practice in developed and developing countries. The PLC typology and integrated PLC framework introduced in this research serve as an innovative product to be applied pragmatically. PLCs represent a long-term sustainable and low-cost approach to teachers' in-service professional development that can be feasibly implemented in low- and middle-income countries and which has the potential to improve teacher practices and student learning.

In addition to guiding policy-makers in the design of PLCs in low-and-middle income countries, the typology proposed in this thesis opens the door for future investigations on which PLC model is best for different settings. The typology and framework have the potential to be complemented and optimized if applied in other contexts, be it within Sub-Saharan Africa, or in other developing regions, where similar structural conditions are found. Another innovative product of this research is the PLC Strength Index, which is a composite index measuring the core characteristics of PLCs as defined by the proposed PLC framework. A variety of actors could benefit from the Index. At the PLC level, PLC participants could apply the index to self-assess the strength of their PLC, identify areas for improvement and monitor progress. The index could also be applied by policy-makers at larger scale to understand how PLCs are functioning and devise strategies to support PLCs accordingly.

As is common in academia, the results from this dissertation will be shared through peer reviewed journals. [39] However, given that these journals are rarely accessed by practitioners and policy makers engaged in in-service teacher professional development, other channels will be explored. In my current role as Technical Advisor at Family Health International 360, which supports design, research and evaluation of education programs in low- and middle-income countries, I am in a great position to also bring the findings to the field. In doing so, I will follow the following strategies:

- Publication of one policy memo shared as an UNU-MERIT memo to be disseminated among the institute's network, which not only includes academics and researchers but also number of practitioners from International Organizations and officials from Sub-Saharan Africa governments.
- Blog posts summarizing the findings and recommendations of this research. I have already published one blog post synthesizing the literature on PLCs at the Research and Evaluation (R&E) Search for Evidence blog. A follow-on blog post on PLCs will be published summarizing the

[39] All four empirical chapters from this dissertation have been submitted as manuscripts to peer reviewed journals and are currently being reviewed by: Comparative Education Review, Comparative Education, Innovations in Education and Teaching, and Teachers and Teaching.

research results. The blog reaches a broad audience, including members of non-profit and multilateral organizations, policy-makers and researchers from different countries.

- Outreach to other countries where FHI360 has education programs to introduce the PLC typology and the thesis research findings with the objective of promoting PLCs within these projects and informing PLC design.

During the PhD, I presented the work on PLCs at the "Evidence Network Conference", held in Pretoria, South Africa, in 2018, as well as at the "Conference for International Education Society (CIES)", held in San Francisco in 2019, which attracts a large audience of education practitioners and academics. Since the presentation at CIES, I have received a number of requests from large nonprofit organizations as well as government representatives who were interested in learning more about the PLC typology and framework to inform the design of PLCs as part of their work. I also published a Blog Post at the R&E Search for Evidence Blog on initial findings from the literature review on PLCs. Finally, the PLC framework has already informed the design of PLCs in an education program implemented by FHI360 in Guatemala.

References

Bashir, S., Lockheed, M., Ninan, E., & Tan, J. P. (2018). *Facing forward: Schooling for learning in Africa*. Washington, DC: World Bank.

Bold, T., Filmer, D., Martin, G., Molina, E., Stacy, B., Rockmore, C., ... & Wane, W. (2017). Enrollment without learning: Teacher effort, knowledge, and skill in primary schools in Africa. *Journal of Economic Perspectives*, *31*(4), 185-204.

Conn, K. M. (2014). *Identifying Effective Education Interventions in Sub-Saharan Africa: A meta-analysis of rigorous impact evaluations* (Doctoral dissertation, Teachers College).

Junaid, M. I., & Maka, F. (2015). In-service teacher education in Sub-Saharan Africa: A synthesis report. *Commonwealth Secretariat. London. Ministry of Education (1996), Educating Our Future: National Policy on Education, Lusaka: Zambia*.

UNESCO Institute for Statistics (UIS). (2016). *Leaving No One Behind: How Far on the Way to Universal Primary and Secondary Education?*. (Policy Paper 27/Fact Sheet 37). Montreal, Canada: Author.

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

Curriculum Vitae

Fernanda was born in Belo Horizonte, Brazil in 1984. She holds a Bachelor's Degree in International Relations from Pontificia Universidade Catolica de Minas Gerais, a Bachelor's Degree in Law from Universidade Federal de Minas Gerais and a Masters of Arts in International Development (Program Evaluation) with honors from American University (Washington D.C.). She decided to engage in development work after completing an internship at the Sustainable Development Department of the Organization of America States (Washington D.C.) while studying Law. Since 2010, Fernanda has worked as a long-term consultant for international organizations, including the Inter-American Development Bank and the Organization of American States, as well as for International NGOs, including Medical Care Development International and Grassroots Soccer. As a consultant, Soares has contributed to the design, evaluation and research of projects in the fields of education, health, youth development and social protection across Latin America and Sub-Saharan Africa.

For the past 5 years Fernanda has served as Technical Advisor within the Research and Evaluation Department in the Global Education, Employment and Engagement Business Unit at FHI 360. In this role, she leads research, monitoring and evaluation (M&E) of education and youth programs in developing countries. Fernanda lives in Michigan, USA, is married to Paul and they have a two-year-old daughter.

UNU-MERIT/MGSoG Dissertation Series

2019

Mira Bierbaum
New Mindsets to Innovate Activation
UNU-MERIT/MGSoG Dissertation Series № 232

Norman Dytianquin
Technology in the Asian Miracle and Crisis Debates: Applications of and Insights from the Field of Influence Approach to Input-Output Analysis
UNU-MERIT/MGSoG Dissertation Series № 231

Nga Le
The implications of health insurance for the labour market and patient satisfaction with medical care in Vietnam
UNU-MERIT/MGSoG Dissertation Series № 230

Jinhyuck Park
Intellectual Property right protection and cross-border R&D investments by multinational enterprises
UNU-MERIT/MGSoG Dissertation Series № 228

Richard de Groot
Show me the Money:
Essays on the Impact of Cash Transfers on Child Nutrition and the Role of Intra-Household Dynamics
UNU-MERIT/MGSoG Dissertation Series № 228

Catie Lott
Diamonds are a Women's Best Friend Broadening Measures of Women's Access to Formal Political Decision-Making
UNU-MERIT/MGSoG Dissertation Series № 227

Ana Cristina Calderon Ramirez
Public Management Reforms
Three stories about public procurement agencification in Latin America
UNU-MERIT/MGSoG Dissertation Series № 226

Camilo Nicanor Carrillo Purin
Teachers' in-service training and student achievement:
The effect of in-service training of Peruvian teachers on student achievement
UNU-MERIT/MGSoG Dissertation Series № 225

Hugo Confraria
Developing scientific capacity in the Global South
UNU-MERIT/MGSoG Dissertation Series № 224

Alison Cathles
Educational Pathways and Skills: Past, Present, and Future
UNU-MERIT/MGSoG Dissertation Series № 223

Ibrahima Sory Kaba
Aggregate Fluctuations and Development: Essays on Macroeconomic Volatility and Economic Growth
UNU-MERIT/MGSoG Dissertation Series № 222

Charlotte Keijser
Firm Participation, Learning and Innovation in Heterogenous Value Chains of IT-enabled Services
UNU-MERIT/MGSoG Dissertation Series № 221

Salih Çevikarslan
Innovation Strategies and Their Implications for Technological Change and Market Outcomes: An Evolutionary Multi-Agent Based Modelling Approach
UNU-MERIT/MGSoG Dissertation Series № 220

Wondimagegn Mesfin Tesfaye
Essays on the Impacts of Climate-Smart Agricultural Innovations on Household Welfare
UNU-MERIT/MGSoG Dissertation Series № 219

Tatevik Poghosyan
How Board Networks Affect Firm Performance and Innovation Incentives in Transition Economies: The Case of Armenia
UNU-MERIT/MGSoG Dissertation Series № 218

Arip Muttaqien
Essays on Inequality and Polarization: Empirical Studies in Developing Asia
UNU-MERIT/MGSoG Dissertation Series № 217

2018

Katrin Marchand
Essays on Forced Migration and Labour Market Participation in Developing Countries
UNU-MERIT/MGSoG Dissertation Series № 216

Ortrun Merkle
The Myth of Gender Neutral Power: Corruption and Gender Norms
UNU-MERIT/MGSoG Dissertation Series № 215

Biljana Meshkovska
Life after Trafficking: (re)integration processes of women that have been trafficked for the purpose of sexual exploitation in Europe
UNU-MERIT/MGSoG Dissertation Series № 214

Vincenzo Vinci
The Relevance of Institutions and People's Preferences for Social Protection
UNU-MERIT/MGSoG Dissertation Series № 213

Silke Heuser
The Effectiveness of Environmental Policies on Reducing Deforestation in the Brazilian Amazon
UNU-MERIT/MGSoG Dissertation Series № 212

Jennifer Waidler
Social Assistance and Remittances and Their Role in the Fight Against Poverty
UNU-MERIT/MGSoG Dissertation Series № 211

Choolwe Muzyamba
The role of community mobilization in the promotion of maternal health of women living with HIV in Zambia
UNU-MERIT/MGSoG Dissertation Series № 210

Juan Carlos A. Castillo Sánchez
Assessing the Role of the Export Sector in Mexican Economic Development,1965-2014
UNU-MERIT/MGSoG Dissertation Series № 209

Tareq Abuelhaj
Food Security Policy Impact Analysis: The Econometrics of Cash and Food Assistance Cost Effectiveness
UNU-MERIT/MGSoG Dissertation Series № 208

Marta Férnandez de Arroyabe Arranz
Essays on M&AS and Innovation
UNU-MERIT/MGSoG Dissertation Series № 207

Clotilde Mahé
Essays on Migration and Occupational Choice
UNU-MERIT/MGSoG Dissertation Series № 206

Simone Sasso
Talent on the move. Essays on Human Capital, Graduate Mobility and Economic Development
UNU-MERIT/MGSoG Dissertation Series № 205

Khaled Walid Rajab
Strategic Planning under Fragility
UNU-MERIT/MGSoG Dissertation Series № 204

Mutinta Hambayi Nseluke
A Tall Order: Improving Child Linear Growth
UNU-MERIT/MGSoG Dissertation Series № 203

Elvis Korku Avenyo
Innovations and Firm Performance in sub-Saharan Africa: Empirical Analyses
UNU-MERIT/MGSoG Dissertation Series № 202

Ni Zhen
Employment Dynamics, Firm Performance and Innovation Persistence in the Context of Differentiated Innovation Types: Evidence from Luxembourg
UNU-MERIT/MGSoG Dissertation Series № 201

Caroline Wehner
Too Scared to Achieve: The Relation Between Neuroticism, Conscientiousness and Socioeconomic Outcomes
UNU-MERIT/MGSoG Dissertation Series № 200

Stefania Innocenti
On Institutional Persistence
UNU-MERIT/MGSoG Dissertation Series № 199

Hassen Abda Wako
Economic Globalization, Institutions and Development: Essays on Aid, Foreign Direct Investment and Trade
UNU-MERIT/MGSoG Dissertation Series № 198

2017

Hans-Erik Edsand
Winds of Change
UNU-MERIT/MGSoG Dissertation Series № 197

Ana Patricia Silva Vara
Redressing the Gender Gap
UNU-MERIT/MGSoG Dissertation Series № 196

Andrés Iván Mideros Mora
Essays on the Economic Effects of Non-contributory Social Protection
UNU-MERIT/MGSoG Dissertation Series № 195

Tobias Broich
New Actors in the Global Economy
UNU-MERIT/MGSoG Dissertation Series № 194

Bernard Nikaj
From No-government to E-government
UNU-MERIT/MGSoG Dissertation Series № 193

Ali Safarnejad
Prioritizing the HIV Response
UNU-MERIT/MGSoG Dissertation Series № 192

Clovis Freire
Diversification and Structural Economic Dynamics
UNU-MERIT/MGSoG Dissertation Series № 191

Michael Verba
Innovation and Knowledge Dynamics: Essays on the Knowledge Economy
UNU-MERIT/MGSoG Dissertation Series № 190

Pui Hang Wong
The Hearts and Minds in Conflict and Peace: The Economics of Counterinsurgency and the Psychology of Reconstruction
UNU-MERIT/MGSoG Dissertation Series № 189

Brenda Yamba
Schooling Despite All Odds: Evidence from Lesotho on Female Child Carers who Stayed in School
UNU-MERIT/MGSoG Dissertation Series № 188

Sheng Zhong
Moving towards An Energy Efficient Future: Essays on Energy Efficiency, Technology and Development
UNU-MERIT/MGSoG Dissertation Series № 187

Julieta Marotta
Access to Justice and Legal Empowerment of Victims of Domestic Violence through Legal Organizations in the City of Buenos Aires: A Qualitative Empirical Legal Study
UNU-MERIT/MGSoG Dissertation Series, № 186

Andrea Franco-Correa
On the Measurement of Multidimensional Poverty as a Policy Tool: Empirical Applications to Chile, Colombia, Ecuador and Peru
UNU-MERIT/MGSoG Dissertation Series, № 185

2016

Yesuf Awel
Insurance for Growth: Empirical Essays on Insurance Demand and Impacts in Africa
UNU-MERIT Dissertation Series, № 108

Tigist Mekonnen Melesse
Grow More Food using Fewer Resources: Agricultural Technology Adoption and Innovation Practices for Inclusive and Sustainable Development
UNU-MERIT Dissertation Series, № 107

Eleni Yitbarek
Getting Ahead or left Behind? Essays on Poverty Dynamics and Social Mobility in Africa
UNU-MERIT Dissertation Series, № 106

Thuy Dieu Nguyen
Firm-Level Theory and Evidence of Corruption
UNU-MERIT Dissertation Series, № 105

Raquel Tsukada Lehman
Essays on Household Production with Labor-Saving Technology
UNU-MERIT Dissertation Series, № 104

Eva Barteková
Multi-Problem Challenges for a Renewable Future: Empirical Studies on Competitive Disadvantages from Electricity Price Differentials and Mineral Supply Risk in an Open Economy
UNU-MERIT Dissertation Series, № 103

Jocelyn Olivari
Entrepreneurial Traits and Innovation: Evidence from Chile
UNU-MERIT Dissertation Series, № 102

Muhammad Shafique
Essays on the role of knowledge, R&D, and Technology-based Firms in the Evolution of Socio-techno-economic System
UNU-MERIT Dissertation Series, № 101

Serdar Türkeli
Governance of Innovation Policy: Empirical Studies on Applied Political Economy by Multi-Methods Analysis
UNU-MERIT Dissertation Series, № 100

Ayokunu Adedokun
Pathways to Sustainable Peace building in Divided Societies: Lessons and Experiences from Mozambique
MGSoG Dissertation Series, № 75

Luiz Rothier Bautzer
Organizing Concurrent Engineering through ICT Platforms Blueprinting Product Lifecycle Management Platforms across Disciplinary Agencies
MGSoG Dissertation Series, № 74

Natalia Popova
Migration in the Periphery of the European Union: Determinants of Successful and Sustainable Labour Market Integration of Return Migrants in Albania, Egypt, Moldova and Tunisia
MGSoG Dissertations Series, № 73

Richard A. Martina
Uncertainty and Resource Constraint in the Small Island Developing States: Essays in Entrepreneurial Cognition
MGSoG Dissertations Series, № 72

Cécile Cherrier
The Expansion of Basic Social Protection in Low-income Countries: An Analysis of Foreign Aid Actors' Role in the Emergence of Social Transfers in Sub-Saharan Africa
MGSoG Dissertations series, № 71

Paul Caldron
The Tacit Bargain in Short-Term Medical Missions: Why U.S. physicians go and what it costs
MGSoG Dissertation Series, № 70

Mahmut Kobal
Customs & Excellence: A Comparative Approach on Administrative and Regulatory Compliance Perspectives of the EU-Turkey Customs Union
MGSoG Dissertation Series, № 69

Craig Loschmann
Essays on Conflict-related Migration and Development in the Case of Afghanistan
MGSoG Dissertations Series, № 68

Andrea Milan
Rural Livelihoods, Location and Vulnerable Environments: Approaches to Migration in Mountain areas of Latin America
MGSoG Dissertation Series, № 67

Farida Lada
On Guarding the Welfare of Clinical Trial Subjects While Promoting Novel Drug Innovation
A Game Theoretical Approach
MGSoG Dissertation Series, № 66

2015

Hibret Belete Maemir
Dissecting Aggregate Productivity: International Integration and Growth with Heterogeneous Firms
UNU-MERIT Dissertation Series, № 96

Giorgio Triulzi
Looking for the Right Path: Technology Dynamics, Inventive Strategies and Catching-up in the Semiconductor Industry
UNU-MERIT Dissertation Series, № 95

Abdul Baseer Qazi
Knowledge flows and networks in the ICT sector: The case of Pakistan
UNU-MERIT Dissertation Series, № 94

Ajay Thutupalli
Technology Paradigm Shifts in Agriculture: Drivers of Sustainability and Catch up
UNU-MERIT Dissertation Series, № 93

Eduardo Urias
Improving access to HIV/AIDS treatment in Brazil: When are Compulsory Licenses effective in Price Negotiations?
UNU-MERIT Dissertation Series, № 92

Francesca Guadagno
Why have so few Countries Industrialised?
UNU-MERIT Dissertation Series, № 91

Daniel Opolot
The Evolution of Beliefs and Strategic Behaviour
UNU-MERIT Dissertation Series, № 90

Alejandro Lavopa
Structural Transformation and Economic Development: Can Development Traps be Avoided
UNU-MERIT Dissertation Series, № 89

Jinjin Zhao
Urban water management reform: The Case of China
UNU-MERIT Dissertation Series, № 88

Simona Vezzoli
Borders, Independence and Post-colonial Ties: the Role of the State in Caribbean Migration
MGSoG Dissertation Series, № 65

Silvia Consuelo Gómez Soler
Civil Conflict and Education: How Does Exposure to Civil Conflict Affect Human Capital Accumulation? Evidence from Standardized Exit Exams in Colombia
MGSoG Dissertation Series, № 64

Paula Nagler
Occupational Choice in the Developing World
MGSoG Dissertation Series, № 63

Jasmin Kientzel
Determinants of Professional Commitment to Environmental Sustainability
MGSoG Dissertation Series, № 62

Mehmet Güney Celbiş
Regional Policies: Convergence, Trade, and the Allocation of Public Capital
MGSoG Dissertation Series, № 61

Florian Henning
Living Up to Standard: Interoperability Governance and Standards Adoption in Government Information Networks
MGSoG Dissertation Series, № 60

Niels P. Groen
The Never-Ending Project Understanding E-Government Project Escalation
MGSoG Dissertation Series, № 59

Derek Copp
Teacher-Based Reactivity to Provincial Large-scale Assessment in Canada
MGSoG Dissertation Series, № 58

Michaella Vanore
Family-Member Migration and the Psychosocial Health Outcomes of Children in Moldova and Georgia
MGSoG Dissertation Series, № 57

Sonja Fransen
The Economic and Social Effects of Remittances and Return Migration in Conflict-Affected Areas: The Case of Burundi
MGSoG Dissertation Series, № 56

Ibrahim Khalil Conteh
The Impact of Floods on Primary School Education in Zambia
MGSoG Dissertation Series, № 55

Richard Bluhm
Growth Dynamics and Development Essays in Applied Econometrics and Political Economy
MGSoG Dissertation Series, № 54

Nevena P. Zhelyazkova
Work-Family Reconciliation and Use of Parental Leave in Luxembourg: Empirical Analysis of Administrative Records
MGSoG Dissertation Series, № 53

2014

Dirk Crass
The Impact of Brands on Innovation and Firm Performance: Empirical Evidence from Germany
UNU-MERIT Dissertation Series, № 87

Samyukta Bhupatiraju
The Geographic Dimensions of Growth and Development
UNU-MERIT Dissertation Series, № 86

François Lafond
The Evolution of Knowledge Systems
UNU-MERIT Dissertation Series, № 85

Annalisa Primi
Promoting Innovation in Latin America: What Countries Have Learned (and What They Have Not) in Designing and Implementing Innovation and Intellectual Property Policies
UNU-MERIT Dissertation Series, № 84

Fatoumata Lamarana Diallo
Evaluation of Meal and Deworming Programs for Primary Schools in Rural Senegal
UNU-MERIT Dissertation Series, № 83

Sachin Kumar Badkas
Metachoice and Metadata: Innovating with Environmental Policy Analysis in Europe
MGSoG Dissertation Series, № 52

Irina S. Burlacu
An Evaluation of Tax-Benefit Systems Impact on the Welfare of Frontier Worker:
The Case of Luxembourg and Belgium
MGSoG Dissertation Series, № 51

Özge Bilgili
Simultaneity in Transnational Migration Research: Links Between Migrants' Host and Home Country Orientation
MGSoG Dissertation Series, № 50

Yulia Privalova Krieger
Reshaping the Big Agenda: Transnational Politics and Domestic Resistance Financial crisis and social protection reform in Bosnia and Herzegovina
MGSoG Dissertation Series, № 49

Marieke van Houte
Moving Back or Moving Forward? Return migration after Conflict
MGSoG Dissertation Series, № 48

Oxana Slobozhan
Global Governance in the Management of Natural Resources: The Case of the Extractive Industries Transparency Initiative (EITI)
MGSoG Dissertation Series, № 47

Luis Bernardo Mejia Guinand
The Changing Role of the Central Planning Offices in Latin America: A Comparative Historical Analysis Perspective (1950-2013)
MGSoG Dissertation Series, № 46

Cheng Boon Ong
Ethnic Segregation in Housing, Schools and Neighbourhoods in the Netherlands
MGSoG Dissertation Series, № 45

Luciana V. Cingolani
Bureaucracies for Development: Oxymoron or Reality? Studies on State Capacity in Challenging Governance Contexts
MGSoG Dissertation Series, № 44

Carlos Cadena Gaitán
Green Politics in Latin American Cities - Sustainable Transport Agendas
MGSoG Dissertation Series, № 43

Katie Kuschminder
Female Return Migration and Reintegration Strategies in Ethiopia
MGSoG Dissertation Series, № 42

Metka Hercog
Highly-Skilled Migration and New Destination Countries
MGSoG Dissertation Series, № 41

Margaret Agaba Rugadya
Can Remittances Influence the Tenure and Quality of Housing in Uganda?
MGSoG Dissertation Series, № 40

Ilire Agimi
New Governance Under Limited Statehood: The Case of Local Government Reform in Kosovo
MGSoG Dissertation Series, № 39

2013

Anant Kamath
Information Sharing through Informal Interaction in Low-Tech Clusters
UNU-MERIT Dissertation Series, № 82

Flavia Pereira de Carvalho
What we talk about when we talk about Brazilian Multinationals: An Investigation on Brazilian FDI, Economic Structure, Innovation and the Relationship between them
UNU-MERIT Dissertation Series, № 81

Jun Hou
Complementarity in Innovation and Development: A Cross-country Comparison
UNU-MERIT Dissertation Series, № 80

Rufin Baghana
Impacts of Government Incentives to R&D, Innovation and Productivity: A Microeconometric Analysis of the Québec Case
UNU-MERIT Dissertation Series, № 79

Lilia I. Stubrin
High-Tech Activities in Emerging Countries: A Network perspective on the Argentinean Biotech Activity
UNU-MERIT/MGSoG Dissertation Series, № 78

Kristine Farla
Empirical Studies on Institutions, Policies and Economic Development
MGSoG Dissertation Series, № 38

Marina Petrovic
Social Assistance and Activation in the Pursuit of Happiness: Shedding New Light on Old Policy Solutions to Social Exclusion
MGSoG Dissertation Series, № 37

Laura Torvinen
Assessing Governance Assessments: The Case of Mozambique: Governance Assessments in the Context of Aid Effectiveness Discourse
MGSoG Dissertation Series, № 36

Biniam Egu Bedasso
Institutional Change in the Long Shadow of Elite: Essays on Institutions, Human Capital and Ethnicity in Developing Countries
MGSoG Dissertation Series, № 35

Sepideh Yousefzadeh Faal Deghati
Childhoods Embargoed: Constructing and Reconstructing Multidimensional Child Poverty in Iran 1984-2009
MGSoG Dissertation Series, № 34

Robert Bauchmüller
Investing in Early Childhood Care and Education: The Impact of Quality on Inequality
MGSoG Dissertation Series, № 33

Martin Rehm
Unified Yet Separated: Empirical Study on the Impact of Hierarchical Positions within Communities of Learning
MGSoG Dissertation Series, № 32

2012

Abdul Waheed
Innovation Determinants and Innovation as a Determinant: Evidence from Developing Countries
UNU-MERIT Dissertation Series, № 77

Bilal Mirza
Energy Poverty and Rural Energy Markets in Pakistan
UNU-MERIT Dissertation Series, № 76

Benjamin Engelstätter
Enterprise Software and Video Games: An Empirical Analysis
UNU-MERIT Dissertation Series, № 75

Fulvia Farinelli
Natural Resources, Innovation and Export Growth: The Wine Industry in Chili and Argentina
UNU-MERIT Dissertation Series

Rodolfo Lauterbach
Innovation in Manufacturing: From Product Variety and Labor Productivity Growth to Economic Development in Chile
UNU-MERIT Dissertation Series

Kirsten Wiebe
Quantitative Assessment of Sustainable Development and Growth in Sub-Saharan Africa
UNU-MERIT Dissertation Series, № 74

Julio Miguel Rosa
Organizational Strategies, Firms' Performance and Spatial Spillovers: The Canadian Case in Research and Development.
UNU-MERIT Dissertation Series, № 73

Johannes Wilhelmus Marie Boels
Joseph Schumpeter, Honderd Jaar Economische Ontwikkeling: Een Historisch-theoretische Beschouwing.
UNU-MERIT Dissertation Series

Dorcas Mbuvi
Utility Reforms and Performance of the Urban Water Sector in Africa
MGSoG Dissertation Series, № 31

Lina Salanauskaite
Distributional Impacts of Public Policies: Essays in Ex-Ante and Ex-Post Evaluation
MGSoG Dissertation Series, № 30

Esther Schüring
To Condition or not – is that the Question?
An Analysis of the Effectiveness of Ex-Ante and Ex-Post Conditionality in Social Cash Transfer Programs
MGSoG Dissertation Series, № 29

Joe Abah
Strong Organisations in Weak States: Atypical Public Sector Performance in Dysfunctional Environments
MGSoG Dissertation Series, № 28

Zina Samih Nimeh
Social Citizenship Rights: Inequality and Exclusion
MGSoG Dissertation Series, № 27

2011

Daniel Vertesy
Interrupted Innovation: Emerging Economies in the Structure of the Global Aerospace Industry
UNU-MERIT Dissertation Series, № 72

Tina Saebi
Successfully Managing Alliance Portfolios: An Alliance Capability View
UNU-MERIT Dissertation Series, № 71

Nora Engel
Tuberculosis in India: A Case of Innovation and Control
UNU-MERIT/MGSoG Dissertation Series, № 70

Evans Mupela
Connectivity and growth in Sub-Saharan Africa: The Role of Communication Satellites
UNU-MERIT Dissertation Series, № 69

Nantawan Kwanjai
Cross Cultural Intelligence amid Intricate Cultural Webs: A Tale of the UnDutchables in the Land of 1002 Smiles
UNU-MERIT Dissertation Series, № 68

Lina Sonne
Innovation in Finance to Finance Innovation: Supporting Pro-poor Entrepreneur-based Innovation
UNU-MERIT Dissertation Series, № 67

Lenka Eisenhamerová
Legitimacy of 'Humanitarian Military Intervention'
MGSoG Dissertation Series, № 26

Sonila Tomini
Informal Payments for Health Care Services in Albania
MGSoG Dissertation Series, № 25

Jinjing Li
Dynamic Microsimulation in Public Policy Evaluation
MGSoG Dissertation Series, № 24

Aziz Atamanov
Rural Nonfarm Employment and International Migration as Alternatives to Agricultural Employment: The Case of Kyrgyzstan
MGSoG Dissertation Series, № 23

Frieda Vandeninden
Poverty Alleviation: Aid and Social Pensions
MGSoG Dissertation Series, № 22

Juliana Nyasha Tirivayi
The Welfare Effects of Integrating AIDS Treatment with Food Transfers: Evidence from Zambia
MGSoG Dissertation Series, № 21

Agnieska Ewa Sowa
Who's Left Behind? Social Dimensions of Health Transition and Utilization of Medical Care in Poland
MGSoG Dissertation Series, № 20

Emmanaouil Sfakianakis
The Role of Private Actors in the Provision of Public Goods with Applications to Infrastructure and Financial Stability
MGSoG Dissertation Series, № 19

Siu Hing Lo
White Collars Green Sleeves: An Inter-organizational Comparison of Determinants of Energy-Related Behaviors among Office Workers
MGSoG Dissertation Series, № 18

Treena Wu
Constraints to Human Capital Investment in Developing Countries: Using the Asian Financial Crisis in Indonesia as a Natural Experiment
MGSoG Dissertation Series, № 17

Henry Espinoza Peña
Impact Evaluation of a Job-Training Programme for Disadvantaged Youths: The Case of Projoven
MGSoG Dissertation Series, № 16

2010

Fernando Santiago
Human Resources Management Practices and Learning for Innovation in Developing Countries: Pharmaceutical Firms in Mexico
UNU-MERIT Dissertation Series, № 66

Zakaria Babutsidze
Essays on Economies with Heterogeneous Interacting Consumers
UNU-MERIT Dissertation Series, № 65

Bertha Vallejo
Learning and Innovation Under Changing Market Conditions: The Auto Parts Industry in Mexico
UNU-MERIT Dissertation Series, № 64

Donatus Ayitey
Technical Change, Competitiveness and Poverty Reduction: A Study of the Ghanaian Apparel Industry
UNU-MERIT Dissertation Series, № 63

Sergey Filippov
Multinational Subsidiary Evolution: Corporate Change in New EU Member States
UNU-MERIT Dissertation Series, № 62

Asel Doranova
Technology Transfer and Learning under the Kyoto Regime: Exploring the Technological Impact of CDM Projects in Developing Countries
UNU-MERIT Dissertation Series, № 61

Florian Tomini
Between Family and Friend: Understanding the Interdependency of Private Transfers
MGSoG Dissertation Series, № 15

Michał Polalowski
The Institutional Transformation of Social Policy in East Central Europe: Poland and Hungary in Comparative and Historical Perspective
MGSoG Dissertation Series, № 14

Maha Ahmed
Defining, Measuring and Addressing Vulnerability: The Case of Post Conflict Environments
MGSoG Dissertation Series, № 13

Pascal Beckers
Local Space and Economic Success: The Role of Spatial Segregation of Migrants in the Netherlands
MGSoG Dissertation Series, № 12

Victor Cebotari
Conflicting Demands in Ethnically Diverse Societies: Ethno political Contention and Identity Values in Europe
MGSoG Dissertation Series, № 11

Dennis Gyllensporre
Competing and Complementary Perspectives on the EU as a Crisis Management Actor: An Examination of the Common Security and Defence Policy through the Lenses of Idealism and Realism
MGSoG Dissertation Series, № 10

Judit Vall Castello
Business Cycle and Policy Effects on Labour Market Transitions of Older and Disabled Workers in Spain
MGSoG Dissertation Series, № 9

Keetie Roelen
False Positives or Hidden Dimensions: The Definition and Measurement of Child Poverty
MGSoG Dissertation Series, № 8

Denisa Maria Sologon
Earning Dynamics in Europe
MGSoG Dissertation Series, № 7

Melissa Siegel
Money and Mobility: Migration and Remittances
MGSoG Dissertation Series, № 6

Jessica S. Hagen-Zanker
Modest Expectations: Causes and Effects of Migration on Migrant Households inSource Countries
MGSoG Dissertation Series, № 5

2009

Alexis Habiyaremye
From Primary Commodity Dependence to Diversification and Growth: Absorptive Capacity and Technological Catch Up in Botswana and Mauritius.
UNU-MERIT Dissertation Series, № 60

Yoseph Getachew
The Role of Public Capital in Economic Development
UNU-MERIT Dissertation Series, № 59

Sandra Leitner
Embodied Technological Change and Patterns of Investment in Austrian Manufacturing
UNU-MERIT Dissertation Series, № 58

Semih Akçomak
The Impact of Social Capital on Economic and Social Outcomes
UNU-MERIT Dissertation Series, № 57

Abraham Garcia
The Role of Demand in Technical Change
UNU-MERIT Dissertation Series, № 56

Saurabh Arora
Coherence in Socio-technical Systems: A Network Perspective on the Innovation Process
UNU-MERIT Dissertation Series, № 55

Mirtha R. Muniz Castillo
Human Development and Autonomy in Project Aid: Experiences from four bilateral projects in Nicaragua and El Salvador
MGSoG Dissertation Series, № 4

Christiane Arndt
Governance Indicators
MGSoG Dissertation Series, № 3

Britta Augsburg
Microfinance: Greater Good or Lesser Evil?
MGSoG Dissertation Series, № 2

2008

Rutger Daems
Medicines for the Developing World
UNU-MERIT Dissertation Series, № 54

Johannes Hanel
Assessing Induced Technology: Sombart's Understanding of Technical Change in the History of Economics
UNU-MERIT Dissertation Series, № 53

Rifka Weehuizen
Mental Capital: the Economic Significance of Mental Health
UNU-MERIT Dissertation Series, № 52

Danielle Cloodt
The Relationship between R&D Partnership Formation, Social Embeddedness and Innovative Performance
UNU-MERIT Dissertation Series, № 51

Sabine Fuss
Sustainable Energy Development under Uncertainty
UNU-MERIT Dissertation Series, № 50

Geranda Notten
Measuring and Managing Poverty Risks
MGSoG Dissertation Series, № 1

2007

Tobias Kronenberg
Reconciling Environmental Conservation with Economic Prosperity: The Feasibility of Double Dividends in the Short and Long Run
UNU-MERIT Dissertation Series, № 49

Viktoria Kravtsova
Assessing the Impact of Foreign Direct Investment in Transition Economies
UNU-MERIT Dissertation Series, № 48

Suhail Sultan
The Competitive Advantage of Small and Medium Sized Enterprises: The Case of Jordan's Natural Stone Industry
UNU-MERIT Dissertation Series, № 47

2006

Bulat Sanditov
Essays on Social Learning and Imitation
UNU-MERIT Dissertation Series, № 46

Mamata Parhi
Dynamics of New Technology Diffusion: A Study of the Indian Automotive Industry
UNU-MERIT Dissertation Series, № 45

Andreas Reinstaller
Social Structures and the Innovation Process: Their Role in the Demand of Firms and Consumers
UNU-MERIT Dissertation Series, № 44

Rose Kiggundu
Innovation systems and Development: The Journey of a Beleaguered Nile Perch Fishery in Uganda
UNU-MERIT Dissertation Series, № 43

Thomas Pogue
The Evolution of Research Collaboration in South African Gold Mining: 1886-1933
UNU-MERIT Dissertation Series, № 42

Geoffrey Gachino
Foreign Direct Investment, Spillovers and Innovation: The Case of Kenyan Manufacturing Industry
UNU-MERIT Dissertation Series, № 41

Önder Nomaler
Technological Change, International Trade and Growth: An Evolutionary, Multi-Agents-Based Modeling Approach
UNU-MERIT Dissertation Series, № 40

2005

Samia Satti Osman Mohamed-Nour
Change and Skill Development in the Arab Gulf Countries
UNU-MERIT Dissertation Series, № 39

Elad Harison
Intellectual Property Rights: Economics and Policy Analysis
UNU-MERIT Dissertation Series, № 38

Daniel Dalohoun
Learning to innovate: agricultural innovation and entrepreneurship: the case of Songhaï farmers in Benin
UNU-MERIT Dissertation Series, № 37

Müge Ozman
Networks, Organizations and Knowledge
UNU-MERIT Dissertation Series, № 36

Bas Straathof
Product Variety and Economic Growth: The Counteracting Effects of Scale and Idiosyncrasy
UNU-MERIT Dissertation Series, № 35

Wilfred Schoenmakers
Knowledge Flows between Multinational Companies: A Patent Data Analysis
UNU-MERIT Dissertation Series, № 34

Myriam Cloodt
Mergers and Acquisitions (M and As) in High-Tech Industries: Measuring the Post-M and A Innovative Performance of Companies
UNU-MERIT Dissertation Series, № 33

2004

Paola Criscuolo
R&D Internationalisation and Knowledge Transfer: Impact on MNEs and their Home Countries
UNU-MERIT Dissertation Series, № 32

Maarten Verkerk
Trust and Power on the Shop Floor
UNU-MERIT Dissertation Series, № 31

Gottfried Leibbrandt
Adoption, Harmonization and Succession of Network Technologies across Countries
UNU-MERIT Dissertation Series, № 30

Mark Sanders
Skill Biased Technical change: Its Origins, the Interaction with the Labour Market and Policy Implications
UNU-MERIT Dissertation Series, № 29

2003

Nadine Roijakkers
Inter-firm Cooperation in High-tech Industries: a Study of R&D Partnerships in Pharmaceutical Biotechnology
UNU-MERIT Dissertation Series, № 28

Viki Sonntag
Speed, Scale and Sustainability
UNU-MERIT Dissertation Series, № 27

Masaru Yarime
From End-of-Pipe Technology to Clean Technology
UNU-MERIT Dissertation Series, № 26

Stéphane Malo
The Combinatorial Chemistry Revolution: Sustaining a Superior Performance Position through Technological Learning
UNU-MERIT Dissertation Series, № 25

2002

Annelies Hogenbirk
Determinants of Inward Foreign Direct Investment: the Case of the Netherlands
UNU-MERIT Dissertation Series, № 24

Bastiaan Johan terWeel
The Computerization of the Labour Market
UNU-MERIT Dissertation Series

2001

John Adeoti
Technology Investment in Pollution Control in Sub-Saharan Africa: The Case of the Nigerian Manufacturing Industry
UNU-MERIT Dissertation Series, № 23

Edward Huizenga
Innovation Management: How Frontrunners Stay Ahead: An Empirical Study on Key Success Factors in the ICT sector
UNU-MERIT Dissertation Series, № 22

2000

Machiel van Dijk
Technological Change and the Dynamics of Industries: Theoretical Issues and Empirical evidence from Dutch Manufacturing
UNU-MERIT Dissertation Series, № 21

1999

Jan Cobbenhagen
Managing Innovation at the Company Level: A Study on Non-Sector-Specific Success Factors
UNU-MERIT Dissertation Series, № 20

Marjolein Caniëls
Regional Growth Differentials: The Impact of Locally Bounded Knowledge Spillovers
UNU-MERIT Dissertation Series, № 19

1998

Aldo Geuna
Resource Allocation and Knowledge production: Studies in the Economics of University Research
UNU-MERIT Dissertation Series, № 18

1996

Reinoud Joosten
Dynamics, Equilibria, and Values
UNU-MERIT Dissertation Series, № 17

Hugo Kruiniger
Investment, R&D, and the Financing Decisions of the Firm
UNU-MERIT Dissertation Series, № 16

1995

Hans van Meijl
Endogenous Technological Change: The Case of Information Technology, Theoretical Considerations and Empirical Results
UNU-MERIT Dissertation Series, № 15

René Kemp
Environmental Policy and Technical Change: A Comparison of the Technological Impact of Policy Instruments
UNU-MERIT Dissertation Series, № 14

Rohini Acharya
The Impact of New Technologies on Economic Growth and Trade: A Case Study of Biotechnology
UNU-MERIT Dissertation Series, № 13

Geert Duysters
The Evolution of Complex Industrial Systems: The Dynamics of Major IT Sectors
UNU-MERIT Dissertation Series, № 12

Marjan Groen
Technology, Work and Organisation: A Study of the Nursing Process in Intensive Care Units
UNU-MERIT Dissertation Series, № 11

1994

Huub Meijers
On the Diffusion of Technologies in a Vintage Framework: Theoretical Considerations and Empirical Results
UNU-MERIT Dissertation Series, № 10

Theon van Dijk
The Limits of Patent Protection: Essays on the Economics of Intellectual Property Rights
UNU-MERIT Dissertation Series, № 9

Hans Voordijk
Naar Integrale Logistiek in Bedrijfsketens: Ontwikkelingen in de Bouw
UNU-MERIT Dissertation Series, № 8

1993

Paul Diederen
Technological Progress in Enterprises and Diffusion of Innovation: Theoretical Reflections and Empirical Evidence
UNU-MERIT Dissertation Series, № 7

Ben Dankbaar
Economic Crisis and Institutional Change: The Crisis of Fordism from the Perspective of the Automobile Industry
UNU-MERIT Dissertation Series, № 6

Hanno Roberts
Accountability and Responsibility: The Influence of Organisation Design on Management Accounting
UNU-MERIT Dissertation Series, № 5

1992

Bart Verspagen
Uneven Growth between Interdependent Economies: An Evolutionary View on Technology Gaps, Trade and Growth
UNU-MERIT Dissertation Series, № 4

Sjoerd Romme
A Self-organization Perspective on Strategy Formation
UNU-MERIT Dissertation Series, № 3

1989

John Spangenberg
Economies of Scale, and Atmosphere in Research Organisations
UNU-MERIT Dissertation Series, № 2

1988

John Hagedoorn
Evolutionary and Heterodox Innovation Analysis: A Study of Industrial and Technological Development in Process Control and Information Technology
UNU-MERIT Dissertation Series, № 1