



Measuring and predicting process quality in Ghanaian pre-primary classrooms using the Teacher Instructional Practices and Processes System (TIPPS)

Sharon Wolf^{a,*}, Mahjabeen Raza^b, Sharon Kim^b, J. Lawrence Aber^b, Jere Behrman^a, Edward Seidman^b

^a University of Pennsylvania, United States

^b New York University, United States



ARTICLE INFO

Article history:

Received 9 March 2017

Received in revised form 30 April 2018

Accepted 17 May 2018

Keywords:

Process quality

Teacher-child interactions

Pre-primary school

Teacher education

Sub-Saharan Africa

School readiness

ABSTRACT

In recent years, there has been an increase in the demand for and supply of early childhood education (ECE) in low- and middle-income countries. There is also growing awareness that unless ECE is of high quality, children may attend school but not learn. There is a large literature on the conceptualization and measurement of ECE quality in the United States that focuses on the nature of teacher-child interactions. Efforts to expand access to high quality ECE in low- and middle-income countries will require similar measurement efforts that are theoretically-grounded and culturally-adapted. This paper assesses the factor structure and concurrent validity of an observational classroom quality tool to assess teacher-child interactions—the Teacher Instructional Practices and Processes System[®] (TIPPS; Seidman et al., 2013)—in Ghanaian pre-primary classrooms. We find evidence of three conceptually distinct but empirically correlated domains of quality: Facilitating Deeper Learning (FDL), Supporting Student Expression (SSE), and Emotional Support and Behavior Management (ESBM). Teachers' schooling level, training in early childhood development, and professional well-being positively predict the three quality domains in different ways. SSE and ESBM predict classroom end-of-the-school-year academic outcomes, and SSE predicts classroom end-of-the-school-year social-emotional outcomes. Implications for the field of international education and global ECE policy and research are discussed.

© 2018 Elsevier Inc. All rights reserved.

1. Introduction

Enrollments in early childhood education (ECE) programs have increased substantially in developing countries over the past several decades (Yoshikawa et al., 2018). In a global initiative to improve children's development and learning, United Nations Sustainable Development Goal 4 on education Target 4.2 aims to expand access to high quality ECE and improve school-readiness outcomes for all children (United Nations, 2015). The target's emphasis on quality is critical, as successful efforts over the last 15 years to increase access to primary education in low- and middle-income countries (LMICs) have not resulted in high learning rates. Conceptualizing and developing culturally-relevant, reliable, and valid metrics of quality is a necessary first step to monitoring and improving it.

Specific features of the classroom context have been shown to meaningfully predict student learning outcomes and trajectories of learning (e.g., Hamre & Pianta, 2001; Mashburn, 2008). Two features of classroom quality are often considered. *Process quality* – considered the driver of child development and learning (Howes, James, & Ritchie, 2003; Pianta, 2005) – refers to the nature of children's daily interactions and experiences in the classroom, with a broad focus on the social, emotional, physical, and instructional aspects of activities and interactions. *Structural quality* refers to regulable resources, such as class size, student-teacher ratio, and teacher training and education (Slot, Leseman, Verhagen, & Mulder, 2015) and are considered important for improving learning outcomes only to the extent that they promote process quality (Phillips, Mekos, Scarr, McCartney, & Abbott-Shim, 2000; Pianta et al., 2005; Vandell, 2004; NICHD Study of Early Child Care, 2002). Yet notably, research is mixed with regard to the consistency of these results across diverse settings, outcomes, and study designs, with many high-quality studies and meta-analyses showing small – or even null – associations between measured classroom quality

* Corresponding author.

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

and child outcomes (Auger, Farkas, Burchinal, Duncan, & Vandell, 2014; Burchinal, 2017).

The vast majority of studies assessing process quality have focused on developed countries. As efforts to expand access to quality education in LMICs increase, establishing metrics that are theoretically-grounded and culturally adapted is crucial. Yet little is known about the conceptual and empirical nature of ECE classroom processes in LMICs generally, and in sub-Saharan Africa (SSA) in particular. In the last two decades researchers have begun to highlight the need for classroom observational instruments that can be used across diverse settings and used in LMICs to gauge elements of classroom quality such as student engagement, effective instruction, and emotional factors that support child development (Schaffer, Nesselrodt, & Springfield, 1994; Venäläinen, 2008). While several instruments have been used in LMICs to assess classroom contexts (Chesterfield, 1997; Stallings & Kaskowitz, 1974; United States Agency for International Development, 2010), most are in the form of frequency-based checklists or time spent on tasks, rather than assessments of process quality.

1.1. Conceptualizing and measuring classroom process quality

Behind the concept of process quality is a long-standing theoretical base on the ways social-settings affect individuals (i.e., through their daily interactions in those settings; Tseng & Seidman, 2007), with a particular focus on classroom settings (Cohen, Raudenbush, & Ball, 2003; Pianta & Hamre, 2009). These frameworks postulate that learning and development rest with students' daily experiences with teachers and peers in the classroom (Seidman & Tseng, 2011), and that these interactions are culturally bound (Stigler, Gallimore, & Hebert, 2000). Whether underlying domains of quality are similar across contexts and cultures is not known, though evidence to date suggests that some quality dimensions are similar across countries. This body of work is primarily limited to the most studied classroom observation tool – the Classroom Assessment Scoring System™ (CLASS; Pianta, La Paro, & Hamre, 2008), and thus it is not clear whether consistencies found across countries are the result of the observation tool itself (see Pastori & Pagani, 2017).

The CLASS organizes teacher-student interactions into three major domains: Emotional Support, Classroom Organization, and Instructional Support (Hamre et al., 2013). These three domains have strong empirical support in high- and a few upper middle-income country contexts (e.g., Hamre et al., 2013). Two recent studies used the CLASS in Latin America. Leyva et al. (2015) found support for the existence of three distinct domains of teacher-child interactions, and Araujo, Carneiro, Cruz-Aguayo, and Schady (2016) used teachers' total CLASS scores to represent Responsive Teaching (as in Hamre, Hatfield, Pianta, & Jamil, 2014). Both studies found that these quality measures predicted children's academic and executive function skills over one school year. Finally, a third study assessed the applicability of the CLASS in kindergarten classrooms in China (Hu, Fan, Gu, & Yang, 2016), and found support for the three-factor structure with acceptable psychometric characteristics.

Other studies, however, have identified psychometric inconsistencies with the original three-factor model (e.g., Pakarinen et al., 2010 in Finland; von Suchodoletz et al., 2014 in Germany). Some have raised critical questions about the international application of standards-based measures to assess ECE quality (Pastori & Pagani, 2017). Interviewing ECE experts and teachers in Italy, Pastori and Pagani (2017) found that while there were elements of continuity in how these teachers perceived quality based on the CLASS dimensions, there were also disagreements about some of the items and key features of the teacher-child relationship that these experts felt were not captured by the tool.

To our knowledge, the CLASS assessment has never been used in SSA and our piloting of the tool suggests it may not be feasible without significant adaptation. For example, the CLASS does not provide guidance on how to use it out of the U.S. or if adaptations are allowed, and its cost is often prohibitive in low-income countries. The cost, coupled with the difficulty of adaptation, is a serious barrier for its use in low-resource contexts. Furthermore, the inferential scoring method, coupled with the phrasing of some of the items, was difficult for local observers to implement with reliability. Finally, classroom assessment tools must be granular and nuanced to be useful for feedback, and according to Pianta (2011), the CLASS dimensions are too coarse for this purpose. This also leads to a need for higher inference on the part of the observer. Measuring classroom process quality in SSA may require a tool developed for the region that is attuned to the context – for example, with a focus on concepts that are particularly relevant when teaching in multi-lingual settings (e.g., tone of voice, language modeling), has sufficient granularity, and is designed with the ease and intent of adaptability by allowing for culturally specific examples to be incorporated.

1.2. The Teacher Instructional Practices and Processes System (TIPPS) – ECE version

The need for a classroom-quality observation tool focusing on process quality is great given its potential as a lever of change to improve education, especially in low-income countries where educating children and youth is a major policy issue (UNESCO, 2008). The Teacher Instructional Practices and Processes System® (TIPPS; Seidman, Raza, Kim & McCoy, 2013) was created to assess classroom process quality in low-income countries. The initial reliability and validity study of the TIPPS was conducted in Ugandan secondary school classrooms (Seidman et al., 2018). As a systematic behavioral observation instrument, the TIPPS has been designed to measure the quality of teaching practices and classroom processes in as granular and nuanced fashion as possible so that they can be fed back to teachers to improve their performance, as well as student academic and social-emotional outcomes.

The development of the ECE version of the TIPPS was, in large part, led by focusing on key quality indicators in the classroom that support child development. These theoretical constructs include the use of structured free-play in the classroom, ECE-focused instructional practices, social-emotional support, classroom management and environment. The measure also includes an “environmental scan” that records the availability of classroom structural supports (e.g., availability of books, desks, and classroom lighting). In the current study, we examine the utility and effectiveness of the pre-primary version of the TIPPS in Ghana.

1.3. Early childhood education in Ghana

Ghana is a lower-middle-income country in West Africa with a population of 26.9 million people (Central Intelligence Agency, 2016). The adult literacy rate is 76.6%, with significant variation across regions. The highest literacy levels (88.5%) are in the Greater Accra Region, with nearly half of the regional population being literate in English and a Ghanaian local language (Ghana Statistical Service, 2012). Nearly one-quarter (24.2%) of the population lives below the national poverty line of approximately \$290 per adult per year (Central Intelligence Agency, 2016). In 2007, Ghana became one of the first countries in SSA to extend two years of pre-primary schooling (i.e., kindergarten for 4 year olds (KG1) and 5 year olds (KG2)) as part of its free universal basic educational system and has among the highest ECE net enrollment rates in Africa at 75% (UNESCO, 2015).

Despite high enrollment rates in schools, educational quality and learning levels remain low. For example, for both Primary 3 and Primary 6, approximately 40% of the students failed to achieve even minimum competency in mathematics and approximately 40% of P3 students failed to achieve minimum competency in English (Ministry of Education, 2014). Recent research suggests that these academic shortcomings may begin early in life, with approximately one-third of three- and four-year-olds in Ghana not meeting basic developmental milestones necessary for success in school, including following directions, working independently, avoiding distraction, and avoiding aggression (McCoy et al., 2016). The relatively high pre-primary enrollment makes ECE a possible target for large-scale improvement of young children's school readiness.

As the demand for ECE has expanded, in particular in peri-urban communities (i.e., towns and settlements that surround large cities), the private sector has grown significantly and helped fill a large gap in communities where public preschools have been slower to open (Bidwell & Watine, 2014; Lewis, 2013). The public and private sectors differ in terms of structural characteristics. For example, there are no requirements for teacher credentials and training in the private sector, while teachers in the public sector are required to have a Diploma in Basic Education obtained from an approved college of education (Asare & Nti, 2014). Public-sector teachers are also classified as civil servants, and thus receive a guaranteed remuneration level as well as job security, while private-sector teachers are generally paid much smaller amounts with no guaranteed job security (Osei, 2006). Understanding if and how observed quality differs in private and public schools is critical to the question of ECE quality in Ghana.

Researchers and policymakers have openly stated the need to reform Ghana's educational system and curriculum, which is based on the colonial model inherited from England (Quist, 2003), if students are to learn the skills necessary to succeed in today's globally competitive economy. In Ghanaian schools, pupils are encouraged to learn by rote memorization and produce lists of facts in compulsory examinations, an approach that neither stimulates creativity nor provides better foundations in academic skills (Agbenyega, 2017). The teaching environment has been described as one characterized by authoritarian leadership styles, overcrowded classrooms, and inadequate teaching resources (Agezo, 2010; Agbenyega, 2017). All public schools and 97% of private schools in the study sample reported using the Ghana Education Service (GES) curriculum. These standards and the corresponding curriculum represent a commitment to comprehensive, child-centered and play-based learning. However, despite reported use of the curriculum, these approaches are often not implemented and the quality of KG education is low across the country (Ghana Education Service, 2012).

1.4. Predictors of process quality

Understanding predictors of process quality is important to inform governments about how to allocate regulable resources. Thus, assessing the convergent and discriminant validity (i.e., the degree to which measures of constructs that theoretically should be related, are in fact related) of variables that predict classroom quality, and how classroom quality predicts child outcomes, are important research questions. While studies have shown that readily-observable teacher characteristics—experience, education, and credentials, among others—explain very little of the differences in teaching quality in the United States (Hanushek & Rivkin, 2012; Rivkin, Hanushek, & Kain, 2005) and in LMICs (Hanushek, 1995), they may be necessary but not sufficient to improving educational systems (Tseng & Seidman, 2007).

One study specifically addressed this question in preschool classrooms in the United States. Drawing on a representative sam-

ple of preschools in six states, Pianta et al. (2005) analyzed the predictors of preschool quality by assessing programmatic features, teacher credentials and ECD training, child-teacher ratio, percentage of students living below the poverty line, and teachers' psychological characteristics. The following factors were correlated with lower observed quality: classrooms with more than 60% of children from homes below the poverty line, teachers lacking formal training in early childhood development, and teachers with less child-centered beliefs. The study also found small but significant positive relations between teachers' psychological well-being and quality, and no relation between structural features of the school building and classroom quality. The findings provide some guidance on what may be important to consider in other contexts.

Similar questions have not yet been explored in SSA. Importantly, teachers in SSA are often ill-equipped for the challenges of teaching in their profession, which threatens their well-being and effectiveness in the classroom. Among the many challenges—which some say have led to a “motivation crisis” (Moon, 2007)—are increasing workloads due to educational reform, low and infrequent compensation, lack of professional recognition and development opportunities, lack of accountability, and lack of voice (Bennell & Akyeampong, 2007; VSO, 2002). Understanding the predictors of high-quality instruction in ECE settings may provide important insights into context-specific levers of change. In this study, we consider four sets of variables in relation to measures of process quality.

1.4.1. Teacher credentials and training

Increasingly in the United States, early childhood advocates are calling for all preschool teachers to have at least a Bachelor's degree (e.g., Barnett, 2004) given evidence that suggests that the schooling level of early childhood education teachers is related to higher quality learning experiences for children. Links have been demonstrated between positive teaching practices and child outcomes with higher levels of education more generally (Howes et al., 2003), and with specific training in early childhood development (Clarke-Stewart, Vandell, Burchinal, O'Brien, & McCartney, 2002). However, a more recent review found largely null or contradictory associations between teachers having at least a BA and high-quality education (Early et al., 2007), concluding that policies focused solely on increasing teacher education are not sufficient to improve quality and learning.

In Ghana, public-sector kindergarten teachers are required to have a minimum of a Diploma in Basic Education obtained from an approved college of education (Asare & Nti, 2014), while there are no minimum requirements for private sector teachers. The certificate in basic education requires two years of coursework, followed by one year of student teaching. Thus, the level of required credentials is less than in high-income countries. It is possible that this training is critical to preparing teachers. But it is equally possible that it is not sufficient to strongly affect teaching quality.

1.4.2. Classroom structural resources

Structural elements of the classroom, such as material resources, class size, and teacher-child ratios, are considered important for improving learning outcomes to the extent that they promote process quality (Seidman & Tseng, 2011; Phillips et al., 2000; Pianta et al., 2005; Vandell, 2004). An influential study released by the NICHD Study of Early Child Care (NICHD Study of Early Child Care, 2002) showed that structural variables of the classroom were associated with key measures of process quality. Children attending programs where caregivers had more training and where child-staff ratios were smaller performed better across a range of measures, and these relations were mediated by childcare process quality.

1.4.3. Teacher professional well-being

Teaching is a stressful profession. Burned-out teachers can create harmful learning environments that are associated with negative outcomes for students. High levels of distress can lead to teacher burnout and deteriorating teacher performance (Osher et al., 2007; Tsouloupas et al., 2010), as well as poorer classroom climate (La Paro et al., 2009). There is growing concern that teachers in lower-income countries, including Ghana, are unmotivated, and that this may partially explain poor teaching performance and student learning outcomes, high rates of turnover and absenteeism, and misconduct (Moon, 2007; Bennell & Akyeampong, 2007). Research focused on child care settings in the U.S. has shown that providers' psychological characteristics were significantly associated with child care quality and interactions with children, with depression and burnout being related to more negative interactions (Hamre & Pianta, 2004). Given the challenges documented in the teaching profession in Ghana (Bennell & Akyeampong, 2007; Osei, 2006), teachers' psychological and professional well-being should be explored as predictors of process quality.

1.4.4. Language of instruction

Ghana is a multi-lingual country, which is a central issue in education. Ghana's National Literacy Accelerated Programme policy (United States Agency for International Development, 2009) mandates that bilingual instruction be used in KG through Primary 3 classrooms, with the proportion of class time taught in English increasing with each year. By Primary 4, students are to be instructed entirely in English. While this policy stipulates that local languages be used for instructions in the early school years, in practice this is often not the case due to various attitudinal, structural and logistical challenges (Ankrah, 2015). With over 44 languages spoken in the country, and with internal migration in urban and peri-urban schools, it is very common for classrooms to have children who speak different native languages. In this case, the teacher is supposed to instruct in the "language of the playground" – that is, the language children use to communicate with each other outside of class (United States Agency for International Development, 2009). Another issue is that in public schools, teachers can be placed in communities across the country and may not be native to the area in which they are teaching. Thus, they may not share a common local language with their students, and in this case teachers often revert to using English.

In this context, we consider whether classroom instruction in local language only, English only, or a mix of English and local language is related to process quality. We consider teachers' proficiency in English (Ghana's official language) and Twi (the most commonly spoken language in the region of this study). In the classrooms in this sample, the majority of schools reported using a mixture of English and local language in their instruction, though there are significant differences in public and private schools. Specifically, 21.7% of private schools reported teaching exclusively in English, none reported teaching in local language only, and 78.3% reported teaching in a mixture of English and local language. In contrast, no public schools reported teaching exclusively in English, 4.3% reported teaching in local language only, while 95.7% reported teaching in a mixture of English and local language.

1.5. The present study

The present study seeks to assess the latent constructs of classroom quality in the ECE version of the TIPPS in Ghanaian classrooms, in addition to the concurrent and discriminant validity of the factors with classroom structural quality, teacher professional well-being, and direct assessments of child school readiness. In assessing the utility of measures of process quality in new settings, the implications for child outcomes is a critical step. Beyond

examining associations with children's outcomes generally, recent research has considered whether certain dimensions of process quality are more strongly related to certain domains of children's development in preschool classrooms (e.g., emotional climate with socio-emotional domain; instructional quality with language and numeracy domains). Specifically, we ask:

1. Are there meaningful factors of classroom quality in a sample of kindergarten classrooms in the Greater Accra Region of Ghana?
2. Given meaningful differences in school and teacher characteristics across public and private schools, is the factor structure invariant across public- and private-sector classrooms? If so, are there differences in classroom quality across the sectors?
3. Is there evidence of concurrent validity for the factors of classroom quality? Specifically, we examine three categories of potential correlates of classroom quality: (1) classroom structural quality (including teacher training and experience and classroom materials); (2) teacher professional well-being, and (3) classroom-levels of child school-readiness skills in the spring (near end-of-the-school-year) controlling for fall (baseline start-of-the-school year) skill levels.

This study contributes to the measurement and conceptualization of classroom process quality in a SSA context. Building on long-standing frameworks of classroom settings and the role of daily interactions (e.g., Tseng & Seidman, 2007; Cohen et al., 2003), we consider underlying concepts of quality in a tool adapted for use in the SSA setting, as well as test several conceptually based and locally relevant predictors of quality. The findings provide insights into potential levers of change to improve educational quality in Ghana specifically, with implications for SSA more generally.

2. Methods

2.1. Participants and procedure

Data for this study come from the baseline of a two-year, three-wave impact evaluation of an in-service teacher training and parental engagement program implemented in six districts in the Greater Accra Region in Ghana (Wolf, Aber & Behrman, 2018). The districts were six of the eight most disadvantaged in the region (out of sixteen total) and ranked between 100 and 187 out of a total of 216 districts in the country (as defined by UNICEF's District League Table, 2015). They included: La Nkwantanang-Madina; Ga Central; Ledzokuku-Krowor; Adenta; Ga East; and Ga South. Schools were identified using the Ghana Education Service EMIS database, which listed all registered schools in the country. Schools were then randomly selected for participation using two levels of stratification: district and public/private schools. In the summer of 2015, 118 public schools and 490 private schools were identified. Ten public schools had no KG classes and were therefore excluded. All public schools were included in the final sample, and private schools were randomly selected in a number proportional to the total number of private schools in each district. The total resulting sample was 108 public schools and 132 private schools.

Every KG classroom in each school was included, and fifteen children (eight from KG1 and seven from KG2) were randomly selected from each school roster to participate in direct assessments. The majority of schools had two KG teachers, though the range was from one to five. If there were more than two KG teachers in the school, two were randomly sampled per school for data collection (one from KG1 and one from KG2). Thirty-seven schools only had one KG teacher, and in this case the one teacher was sampled. In the fall, several teachers had not yet been posted to their classrooms; only 420 of the 443 classrooms had teachers present when data

collection took place. Nearly all of those teachers (97.9%) agreed to be videotaped and participate in the survey, which was conducted by trained enumerators. However, due to the theft of three tablets and two netbook computers, 94 classroom videos were lost. Thus, the total sample size for the classroom observations was 317 teachers/classrooms. While theft is exogenous, we assessed for potential bias in the loss of classroom videos by assessing statistical differences in each predictor variable. There were no statistically significant differences between the two groups (results available by request to first author).

All teachers were videotaped teaching a lesson in their classroom for 30–60 min in the first half of the six-hour school day in the first term of the school year in September or October of 2015. Videos were recorded by trained enumerators who also administered surveys to the teachers. The content of the lessons varied, and were reported by teachers to include: creative activities (14.5%); environmental studies (10.1%); literacy and language (35.7%); mathematics (33.8%); movement and dance/physical education (3.0%); and “other” (2.8%).

2.2. Measures

2.2.1. Teacher instructional practices and processes system – ECE version

The TIPPS-ECE version was adapted from the primary and secondary versions of the TIPPS (Seidman et al., 2013). Adaptations were made in a number of arenas. The scopes of certain dimensions were modified to more accurately reflect processes appropriate to the developmental level. For example, the cooperative-learning dimension was changed to focus more on the basic skills necessary (i.e., learning how to share and play together) for cooperative learning in ECE. Additionally, the complex-thinking/problem-solving dimension was reframed to reflect how the teacher might help children use language to process information and promote problem solving (share reasoning aloud). Other dimensions were added to reflect aspects of the classroom that were unique to ECE, such as children’s learning through the use of free play, as well as a larger emphasis on behavioral management (i.e., how misbehavior is managed) and classroom management (i.e., creation and reinforcement of consistent classroom routines). All examples in the codebook were changed accordingly. The changes were aided by familiarity with the ECE literature, as well as focus groups, pilot testing, and feedback from local ECE specialists.

Each of the 19 TIPPS concepts¹ is broken down into three areas in the training manual: (i) *importance of concept*, (ii) *behavioral indicators*, and (iii) *concrete examples*. For example, the importance of the first TIPPS ECE dimension, *Teacher supports children’s development through the use of free playtime*, is explained in the importance of concept by stating its role in contributing to a quality learning environment. A behavioral indicator outlines a possible manifestation of this concept. An example in this case could be to examine whether or not the teacher allows children to structure their activities without imposing any formal way to play. Finally, concrete examples help illustrate each of the four scale points to illustrate the “degree” to which the concept is present in the classroom.

To facilitate objective observations and ratings, the TIPPS employs a “structured alternative format” (adapted from the work of Harter, 1982). In this structured format, both visually and conceptually, the rater is presented with a concept to make a dichotomous decision of whether the concept is absent or present

in the classroom (Column A or Column B). Once this decision is made, the rater is asked to make a second dichotomous rating by deciding if the statement they chose is “somewhat accurate” or “very accurate.” The manual was designed to be short, focusing exclusively on a guiding rationale for each core concept, its behavioral indicators, examples of each teacher practice and classroom process, and four-point rating format.

After piloting the instrument in classrooms and conducting cognitive interviews with KG teachers about the tool, only minor adaptations were made for use in Ghana (e.g., referring to pupils as children, as is common in Ghanaian KG settings). The concepts and examples in the manual were deemed appropriate.

2.2.1.1. Coding protocol and rater training process. All videos were watched and coded with the TIPPS (starting at minute five of the video through minute 30). All raters were Ghanaian nationals and had at least a Bachelor’s degree. They were required to attend a 5-day training focused on introducing the TIPPS, bias awareness and minimization strategies, and practice classroom videos. Upon completing the training, each rater had to meet or exceed three calibration criteria within three attempts to be certified as a TIPPS observer. Master Codes for the calibration videos were established by two of the TIPPS creators and the projects’ principal investigator, who watched each video and established agreement for the respective codes.

The criteria were as follows: (1) 60% or more of the individual trainee’s codes had to perfectly match the Master Codes for all calibration segments; (2) an individual trainee’s rating codes had to match or be within one rating category of the Master Code, for at least 90% of all his/her ratings; and (3) across all concepts, a trainee could only deviate from the Master Coded Category (i.e., Column A or B) on one item. In particular, a high rate of dichotomous mismatches would indicate a lack of understanding of the TIPPS concepts and guiding rationales. The calibration criteria not only assessed agreement, but also the degree of deviation from master codes – both important aspects given that there are only four scale points and that understanding of the concept is critical for precise coding. Of the eleven trainees who attended the training, six passed the calibration. Throughout the coding process, the six raters participated in weekly meetings with the project’s principal investigator to discuss questions and issues regarding the coding process and to reduce rater drift.

2.2.1.2. Rater reliability. To assess inter-rater reliability, 15% of videos were coded by two additional raters (three total). We calculated the intraclass correlation coefficient (ICC) of the final scores to assess how the partition of variance in scores breaks down into differences in individual raters and how much variance is shared across raters. On average across items, 71.1% of the variance was shared across raters, denoting acceptable inter-rater reliability.

2.2.2. Validation measures

2.2.2.1. Teacher training. Teacher training backgrounds were collected in an administered survey. We assessed teacher’s *schooling level* (coded as 1 = primary or Junior High School, 2 = Secondary High School (SHS) only, 3 = SHS+ but no college degree; 4 = BA degree; 5 = MA or PhD; $M = 2.79$, $SD = 0.87$); *total years working as a kindergarten teacher* ($M = 6.7$, $SD = 7.0$, $range = 0–40$), and whether the teacher had specific *training in ECD* (66.2% yes). We also included a measure of knowledge of pre-literacy skills (i.e., *reading knowledge*) through the percentage of correct answers in an 8-item task to assess teachers’ basic knowledge level ($M = 54.8\%$ correct, $SD = 17.3$).

2.2.2.2. Teacher professional well-being. Teachers answered an administered survey in English. All items came from existing scales

¹ Here, the term ‘concept’ is used interchangeably with the term ‘dimension’. While ‘dimension’ is more consistent with the literature around the measurement of classroom quality, we also retain the use of concept, especially in training raters, as a more broadly accessible term.

and factors were derived through exploratory factor analyses. We measured four constructs pertaining to teachers' well-being. *Personal accomplishment* was measured by eight items from the Maslach et al. (1996) Burnout Inventory ($\alpha = .64$). Sample items included "I deal very effectively with the problems of my school children" and "I have accomplished many worthwhile things in this job". *Motivation* was measured by five items ($\alpha = .77$) adapted from Bennell and Akyeampong (2007) as reported in Torrente et al. (2012). Sample items included "I am motivated to help children develop well socially (i.e., behave well, get along with peers, cooperate)" and "I am motivated to help children learn math." *Job satisfaction* was measured by four items ($\alpha = .72$) adapted from Bennell and Akyeampong (2007) by Torrente et al. (2012). Sample items included "I am satisfied with my job at this school" and "I want to transfer to another school (reverse coded)." Finally, *burnout* was measured through 11 items ($\alpha = .85$) from the Maslach et al. (1996) Burnout Inventory. Sample items included "I feel I am working too hard on my job" and "I feel emotionally drained from my work." Items within a scale were averaged. Scales were all normally distributed.

All items were pilot tested. First, we conducted five cognitive interviews (Willis, 2005) with teachers to assess whether they understood the questions consistently across subjects and in the ways intended (Collins, 2003). We then piloted the survey by administering it to 20 teachers through one-on-one surveys. From both of these exercises, we concluded that these items were suitable for use in this sample. All items had been used in previous research with teachers in SSA (Wolf, Torrente et al., 2015; Wolf, Aber et al., 2015).

2.2.2.3. Classroom resource index. The TIPPS classroom environmental scan was an observational supplement to gauge a classroom's condition and materials. Its purpose was to aid in understanding some of the classroom's basic assets, needs, and the surroundings in which the class was operating. Using the environmental scan, each class was scored based on the number of structural resources available (nine total), with higher scores indicating more resources. Specifically, classes were scored on: number of teachers in the classroom ($0 = 0$ (i.e., only teacher's aide), $0.33 =$ one teacher, $0.67 =$ two teachers, $1 =$ three teachers); class size ($0 =$ greater than 25 students, $1 =$ fewer than 25); classroom physical structure ($0 =$ open/outdoors, $0.5 =$ covered, open sides, $1 =$ covered, four walls); seating furniture for students ($0 =$ bare floor or mats, $0.5 =$ seats but no tables, $1 =$ seats with tables); books in the classroom ($0 =$ no books, $1 =$ at least some books); if there was adequate writing space for all students ($0 =$ no, $1 =$ yes); if there was a quiet learning environment ($0 =$ noise affects communication; $1 =$ noise does not affect communication); posters/charts in the classroom ($0 =$ no, $1 =$ yes); and adequate lighting in the classroom ($0 =$ no, $1 =$ yes). The index was normally distributed ($M = 7.0$, $SD = 1.4$, range = 0–9).

2.2.2.4. Language of instruction and proficiency. Classrooms were rated by assessors on the language of instruction: whether the lesson was delivered in English only ($N = 206$, 65.0%), local language only ($N = 5$, 1.6%), or a mix of English and local language ($N = 106$, 33.4%). Second, we used teachers' self-reports whether they were proficient in speaking English (40.8%) and Twi (50.6%). Of the remaining teachers, 49.4% reported speaking English at an intermediate level and 9.9% at a basic level. For Twi, 29.6% reported speaking at an intermediate level, 15.6% at basic level, and 4.2% reported speaking none. We dichotomized both variables to indicate proficient (=1) versus not proficient (=0). Notably, all enumerators were native Ghanaians and were able to translate items into local languages if teachers did not understand the English version. Spe-

cific wording for translating several of the more difficult items was agreed upon by all enumerators during training.

2.2.2.5. Classroom school-readiness skills. Children's school readiness was assessed directly using the International Development and Early Learning Assessment (IDELA; Pisani et al., 2015) in the fall (2015) and spring (2016) of the school year. The IDELA is a direct measure of child learning and development that has been used in over 30 countries. The tool takes approximately 35 min to administer and was derived from several commonly used assessments of school readiness, including the Early Development Instrument (EDI; Janus et al., 2007), the Ages and Stage Questionnaire (ASQ; Squires & Bricker, 2009), the Malawi Development Assessment Tool (MDAT; Gladstone et al., 2010), and the East Asia Pacific-Early Child Development Scales (EAP-ECDS; Rao et al., 2014). (Further details on the development and implementation of the IDELA can be found in Pisani et al., 2015). The factor structure of the IDELA has been confirmed across five countries in a previous study (Halpin et al., 2018). The current protocol was piloted and adapted to the Ghanaian context in which this study occurred.

The assessment included items that covered five domains of children's development, specifically Emergent Numeracy, Emergent Literacy, Gross and Fine Motor Skills, Social-Emotional Development, and Executive Function (specifically, inhibitory control and working memory). In more detail, *Emergent Numeracy* was captured using 39 items grouped into eight subtasks and covers constructs of number knowledge, basic addition and subtraction, one-to-one correspondence, shape identification, sorting abilities based on color and shape, size and length differentiation, and completion of a simple puzzle ($\alpha = .72$). *Emergent Literacy* was measured using 38 items grouped into six subtasks, and covered constructs of print awareness, letter knowledge, phonological awareness, oral comprehension, emergent writing, and expressive vocabulary ($\alpha = .74$). *Social-emotional Development* was measured using 14 items grouped into five subtasks, and covered constructs of self-awareness, emotion identification, perspective taking and empathy, friendship, and conflict/problem solving ($\alpha = .69$). *Motor Development* was measured using nine items grouped into three subtasks assessing fine motor skills ($\alpha = .83$). Finally, *Executive Function* was assessed through ten items grouped into two subtasks focused on working memory (i.e., digit span) and impulse control ($\alpha = .83$).

Children's scores in each of the five domains were calculated through averaging scores on each of the sub-domain tasks. Scores in each of the five domains were then averaged to create a "school readiness" score for each child. Scores for each domain, as well as total school-readiness scores, were aggregated to the classroom-level to investigate processes at the level of the teacher/classroom and given the relatively high mobility of students across classrooms and schools in this context. Intra-class correlations for the outcomes varied, with 19.9–37.8% of the variance in child outcomes explained at the level of the classroom.

2.3. Analytic plan

The four steps of our analyses are described below. All analyses adjusted standard error estimates for the nesting of teachers ($N = 317$) within schools ($N = 179$). By design, there were no more than 2 teachers sampled per school ($M = 1.77$).

2.3.1. Preliminary analysis of TIPPS

Four items were dropped due to low variation: Item 1 (*free play*; 100% of classrooms had a rating of 1), Item 2 (*cooperative play*; 94.0% of classrooms had a rating of 1), Item 15 (*gender equity*; 98.4% of the classrooms had a rating of 2), and Item 16 (*non-gender based equity*; 98.7% of classrooms had a rating of 2). The remaining fifteen items

were used in all subsequent analyses (descriptive statistics shown in Appendix A, Table A1).

2.3.2. Exploratory and confirmatory factor analysis

We randomly divided the sample of classrooms into two halves: an exploratory sample ($N=159$) and a confirmatory sample ($N=158$). The purpose of the exploratory sample was to allow for multiple variations of the sixteen items on initial models to be fitted in order to arrive at a proposed model. The purpose of the confirmatory sample was to ensure that the proposed model demonstrated out-of-sample generalizability. The N:k ratio of approximately 10:1 for both the exploratory and confirmatory models provided sufficient power to run the two analyses (Ford, MacCallum, & Tait, 1986).

Analyses were conducted using MPlus V6.12. We examined factor solutions ranging from one to four factors in the exploratory analysis using varimax rotation. In selecting a model, we considered (1) goodness-of-fit statistics and (2) conceptual meaning of the factors. The selected model was examined in a confirmatory factor analysis using the confirmatory sample.

2.3.3. Measurement invariance

The second step built on the model finalized in the first step by formally testing the measurement invariance of items contributing to each factor for two groups using the full sample: public ($N=165$) and private ($N=152$) schools. Measurement invariance is an important step to establish whether differences among subgroups' observed scores can be explained only in terms of differences on the constructs/factors of interest. If measurement invariance does not hold, differences among subgroups' observed scores can be explained by additional features of the groups, and these additional features are generally treated as sources of measurement bias, or as a theoretically meaningful difference, depending on the a priori underlying model.

There are three commonly assessed types of measurement invariance, which form a nested hierarchy representing increasing levels of invariance (e.g., Millsap, 2012; Teresi, 2006). The first two levels, dimensional and configural invariance, establish that the number of factors and general pattern of item loadings are the same across groups. We began by testing for configural invariance. The second level, metric (or pattern) invariance, establishes that the value of each item's loading on the factor is the same across groups. This allows for comparisons of the (co-) variances of the factors across groups. The third level of invariance, scalar (or strong factorial) invariance, establishes that the value of the intercept or threshold for each item is equivalent across groups. This allows for comparisons of the means of the factors across groups.

2.3.4. Concurrent and predictive validity

The final analysis tested for concurrent and predictive validity. To test concurrent validity, we assessed how teacher characteristics related to classroom scores on each of the three factors. This test of validity allowed for the examination of whether characteristics expected to be related to teaching quality are, in fact, related. First, we computed bivariate correlations between each predictor and each factor (see Table 3). Second, we used latent regression analysis to assess the relationship between predictors with each factor tested in three separate models using a structural equation model (SEM) framework (Model 1). These results can be interpreted as the relationship between each predictor and factor, controlling for the relationships between other predictors with that factor. Finally, we tested for predictive validity of the classroom quality factors. We added to the latent regression analysis spring classroom-level child outcomes in each of the five domains, controlling for classroom-level child outcomes in the fall (Model 2). These results can be interpreted as the relations between each predictor and domain

of quality, and the subsequent relations between spring classroom outcomes controlling for fall classroom outcomes. All latent regression analyses adjusted standard error estimates for clustering of classrooms within schools.

3. Results

Descriptive statistics for all independent variables are presented in Table 1, broken down by public and private schools. Almost all (98%) teachers were female. On average, teachers had 6.7 years of experience teaching kindergarten. There were large differences in teachers' schooling level by public and private status, with nearly all (95%) public school teachers having education beyond secondary school compared to less than one-quarter (24%) of private school teachers. Similarly, public school teachers were more likely to have formal training in early childhood development compared to private school teachers (73% versus 58%). Notably, public- and private-sector classrooms had similar levels of material resources. The average class size was 25 ($SD=11$, range=3–70), with public school classrooms being larger than private school classrooms (means of 28 vs. 21, respectively, $F=32.6$, $p<.001$).

3.1. Factor analysis

3.1.1. Exploratory factor analysis

A three-factor solution was chosen based on a combination of goodness-of-fit statistics and interpretive, conceptual analysis of results. The one-factor solution was not chosen because of poor model-fit statistics ($\chi^2(104)=467.39$, $RMSEA=.148$). The two-factor solution had acceptable fit ($\chi^2(89)=168.80$, $RMSEA=.079$) but was not selected, primarily because the conceptual interpretation of the two factors was not clear; the items on both factors contained a mix of content in terms of instructional and emotional dimensions. In the three-factor solution ($\chi^2(75)=112.24$, $RMSEA=.056$)—the selected model—a third factor representing emotional support was distinct. This allowed for easier interpretation of instructional vs emotional components of the classroom environment. We used a 0.40 factor loading as a cutoff criterion for including items on each factor (Hair, Tatham, Anderson, & Black, 1998; Tabachnick & Fidell, 2007). One item—Item 5 (instructional materials)—did not load on any factor above .40 and was dropped from the analysis. No items loaded above .40 on more than one factor.

We labeled the three factors as *Facilitating Deeper Learning* (factor 1), *Supporting Student Expression* (factor 2), and *Emotional Support and Behavior Management* (factor 3; Appendix A, Table A2). While the four-factor solution had slightly better goodness-of-fit statistics, the solution had two factors with only two items, and thus was not selected for sake of parsimony. We also assessed the goodness of fit for a bifactor model to assess an over-arching quality factor (similar to that of *Responsive Teaching*; Hamre et al., 2014). This model did not fit the data well. Thus, a three-factor model was selected.

3.1.2. Confirmatory factor analysis

Using the confirmatory sample, we replicated the three-factor model using a confirmatory factor analysis. The results are shown in Fig. 1. All loadings ranged from 0.53 to 0.85. The confirmatory model had adequate goodness-of-fit statistics ($\chi^2(df)=154.19(87)$, $RMSEA=.070$, $CFI=.946$, $TLI=.934$). The correlation between *Facilitating Deeper Learning* (FDL) with *Emotional Support and Behavior Management* (ESBM) was $r=.67$; FDL with *Supporting Student Expression* (SSE) was $r=.26$; and ESBM with SSE was $r=.72$.

Table 1
Descriptive Statistics of Independent Variables, by Public and Private Sector Status.

	Public Mean/SD	Private	Min	Max		
Total years as KG teacher	6.99	7.31	6.34	6.54	0	40
Schooling level (%)						
Junior high school	3.0		9.0		0	1
Secondary high school	1.8		66.9		0	1
>Secondary high school, no college degree	54.6		23.5		0	1
BA degree	39.4		0.7		0	1
MA or PhD	1.2		0.0		0	1
Training in ECD (%)	73.3		58.4		0	1
Reading knowledge ^a	0.564	0.173	0.529	0.172	0	1
Personal accomplishment	5.88	0.96	5.78	0.97	1.75	7
Burnout	2.13	0.99	2.04	0.93	1	7
Motivation	4.53	0.72	4.74	0.41	1	5
Job satisfaction	1.82	0.63	1.91	0.67	1	3.83
Classroom resource index	7.04	1.37	7.04	1.10	0	9
Sample size (total = 317)	165		152			

^a Scored as a percent correct from 0 to 1.

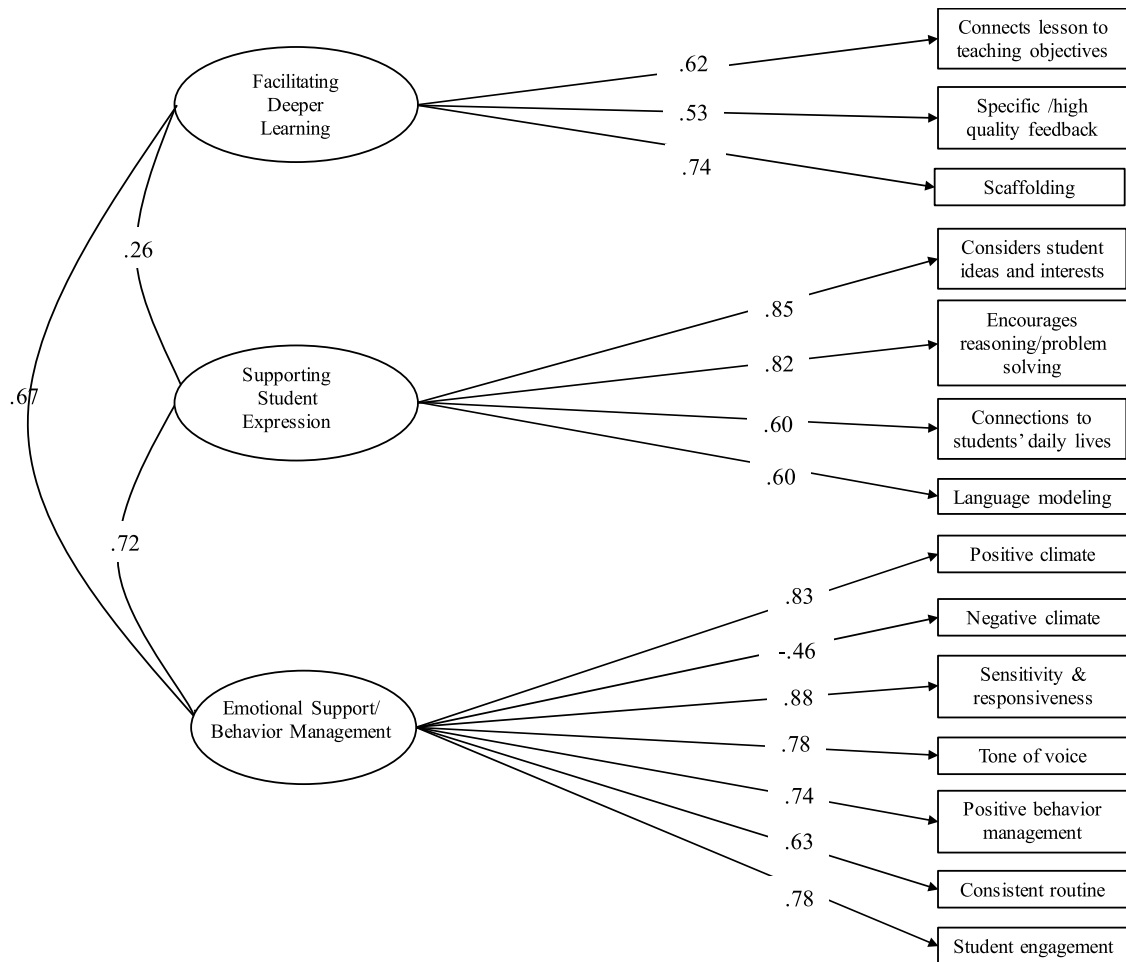


Fig. 1. Final Confirmatory Factor Model.

Table 2
Model Fit Statistics for Measurement Invariance Tests for Public versus Private Sector Classrooms.

	Configural invariance	Metric invariance	Scalar invariance
Chi-square (df) Contributions:	374.17 (204)	361.55 (216)	378.21 (231)
Public	206.76	198.55	205.51
Private	167.41	163.00	172.70
CFI	.936	.945	.944
TLI	.934	.947	.949
RMSEA	.071	.066	.064
χ^2 change (df)		10.62 (12)	31.33 (27)
χ^2 change p-value		.562	.258

3.2. Measurement invariance across public- and private-sector classrooms

The next step of our analysis addressed the question of whether the factor structure was invariant across and private- and public-sector schools/classrooms. The three-factor model depicted in Fig. 1 was fitted to each group. The goodness-of-fit of the comparisons supported the conclusion of scalar (strong factorial) measurement invariance (see Table 2).

Having established measurement invariance across the groups, we then assessed the mean differences, in standard deviation units, for each factor. We found no differences when comparing means scores in private versus public schools in Facilitating Deeper Learning ($SD\ diff = 0.05, p = .773$) or in levels of Supporting Student Expression ($SD\ diff = 0.03, p = .804$). We did find significant differences in levels of Emotional Support and Behavior Management, with private schools displaying lower scores on this factor of over one-third of a standard deviation compared to public schools ($SD\ diff = -0.35, p = .012$).

3.3. Concurrent and predictive validity

Table 3 displays the correlations among predictor variables and the three classroom quality factors. Years of experience as a kindergarten teacher was not correlated with any domain of classroom quality. Schooling level, training in ECD, and reading knowledge were all correlated in expected directions with Supporting Student Expression (SSE), showing small to moderate associations, but not with Facilitating Deeper Learning (FDL) or Emotional Support and Behavior Management (ESBM). Dimensions of professional well-being correlated with classroom quality in the expected directions, with the correlations small in size. Classroom resources and classroom-level child school-readiness skills had small to moderate correlations with quality. Finally, SSE was positively correlated with teachers use of local language in the lesson delivery and negatively with use of English only, with correlations small in size; and teachers' proficiency in Twi had a small positive correlation with ESBM.

In predicting classroom quality factor scores using latent regression analysis, we conducted three regression analyses (one for each factor). The results are summarized in Table 4. Number of years as a kindergarten teacher was not related to any factor. Schooling level and training in ECD were both related to higher scores of SSE ($\beta = .25, p < .001$ and $\beta = .17, p < .01$).

Some dimensions of teachers' professional well-being were also related to classroom quality in the expected directions, though the patterns were not consistent across factors. Personal accomplishment was associated with higher levels of FDL ($\beta = .24, p < .001$) and SSE ($\beta = .14, p < .05$). Burnout was negatively associated with SSE ($\beta = -.22, p < .01$). The classroom resource index was positively associated with ESBM ($\beta = 0.26, p < .001$) but not SSE or FDL.

When considering language, language of instruction (English, local language, or English and local language mix) did not predict classroom quality. Furthermore, teachers' proficiency in English and Twi did not significantly predict classroom quality, with one exception: teachers' proficiency in English marginally significantly positively predicted ESBM ($\beta = 0.12, p < 0.10$).

In the final models (Models 2; Table 4), child outcomes across five domains from the fall and spring assessments were averaged to the level of the classroom and added to Model 1, with classroom quality factor scores predicting spring outcomes controlling for fall outcomes (i.e., predictors \rightarrow classroom quality factors \rightarrow spring classroom-level child outcomes, controlling for fall classroom-level child outcomes). FDL marginally statistically significantly predict higher spring early numeracy skills ($\beta = 0.07, p < 0.10$). SSE statistically significantly predicted higher early literacy scores

($\beta = 0.17, p < 0.01$) and social-emotional score ($\beta = 0.13, p < 0.05$), and marginally significantly predicted higher motor skills ($\beta = 0.10, p < 0.10$). Finally, ESBM predicted higher early numeracy scores ($\beta = 0.08, p < 0.05$) and marginally significantly predicted early literacy scores ($\beta = .08, p < .10$).

4. Discussion

We assessed the factor structure of ECE classroom process quality and its correlates in Ghana, as well as the utility of process quality in predicting child academic and behavioral outcomes. In a period of growing demand for the provision high quality ECE, as well as for monitoring systems and metrics of quality, we provide evidence that the Teacher Instructional Practices and Processes System© (TIPPS; Seidman et al., 2013; 2018) is one tool that may be suitable for this purpose. We assessed quality through teacher-child interactions, given its' critical role in promoting child learning and development in other contexts (Cohen et al., 2003), and showed that this is a useful lens through which to assess classrooms in SSA. While many interventions aim to change specific instructional strategies and curricula that are thought to be effective, there is also a need for clearer articulations and validation of higher-level theories of effective teaching that include process quality within and across countries. This study is one of the first to attempt this work in pre-primary classrooms in SSA.

We found three domains of classroom quality in this sample of public and private schools. We discuss them in light of the current evidence of ECE classroom quality, primarily in the United States, to draw out similarities and differences in two different educational systems. The first factor – Facilitating Deeper Learning (FDL) – involves instructional support strategies that teachers use to encourage learning (e.g., scaffolding, providing high quality feedback), and may be similar to the factor of Instructional Support found in ECE classrooms in the United States (Teachstone Training LLC, 2014). Notably, FDL did not predict child learning outcomes in this context. This may be due to teachers' lack of mastery of the academic content. For example, in a brief pre-literacy knowledge scale (assessing knowledge about, e.g., syllables, speech sounds, rhyming words), the average score in this sample was 54%. This points to a need for teachers' pre-service training to focus more on content and curricular knowledge, in addition to pedagogy. Teachers' proficiencies in the language of instruction may also be a barrier to their ability to facilitate deeper learning. Only 40% of teachers reported proficiency in English, and only half reported proficiency in Twi, the two dominant languages of instruction in the districts in this study.

The second factor – Emotional Support and Behavior Management (ESBM) – involves both concepts related to supporting students emotionally (e.g., sensitivity and responsiveness, tone of voice) and positive behavior management strategies (e.g., providing a consistent routine). The existing literature has found emotional support and classroom organization (including behavior management) as two distinct domains of classroom quality (Leyva et al., 2015; Teachstone Training LLC, 2014). Our findings suggested that behavior management and emotional climate were closely connected in Ghanaian classrooms. Notably, this was the only factor in which we found a meaningful difference between public and private schools, with lower quality found in private-sector classrooms. This suggests that high parental demand for "rigorous" education in Ghana ECE (Bidwell & Watine, 2014) may affect the private sector more strongly. Alternatively, it is possible that this domain of instructional practice was most sensitive to the differences in teacher training and credential requirements in public and private schools.

Table 3
Bi-variate Correlations among Study Variables.

	FDL	SSE	ESBM	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1 Total years as KG teacher	-.04	.04	.01														
2 Schooling level	.08	.24	.06	-.08													
3 Training in ECD (yes/no)	.08	.18	.10	.35	.13												
4 Reading knowledge	.04	.11	.04	-.14	.09	.00											
5 Personal accomplishment	.20	.16	.09	-.08	.13	.07	.03										
6 Burnout	-.08	-.09	.05	-.09	.08	-.05	.16	-.05									
7 Motivation	.00	-.05	.11	-.06	-.08	-.04	.00	.17	-.12								
8 Job satisfaction	.19	.01	.09	-.10	.02	.01	.12	-.03	.32	-.13							
9 Classroom resource index	.06	.11	.19	.04	.02	-.02	.04	-.06	-.07	.01	-.02						
10 Lesson delivered in English	.03	-.10	.02	-.01	-.43	-.12	-.08	-.06	.02	.12	.01	-.16					
11 Lesson delivered in local language	.00	.13	.02	.00	.06	.04	.15	.03	-.06	.01	.00	.06	-.17				
12 Lesson delivered in mix of English and local language	-.05	.19	-.03	.08	.41	.11	.04	.06	.00	-.12	-.01	.15	-.97	-.09			
13 Teacher is proficient in English	.13	.16	.11	-.09	.34	.07	.10	.17	.00	-.08	.07	.08	-.15	.05	.13		
14 Teacher is proficient in Twi	.03	-.01	.15	.03	.05	.09	.00	.08	-.09	.04	-.06	.09	-.06	-.08	.09	.34	
15 Average classroom school readiness (total)	.20	.08	.20	.02	.03	.03	.08	.03	.09	-.01	.05	-.07	.08	.03	-.09	.04	.01

Note: Bold numbers indicate correlation is statistically significant at $p < .05$. Correlations with FDL, SSE, and ESBM derived from respective factor scores. FDL = Facilitating Deeper Learning; SSE = Supporting Student Expression; ESBM = Emotional Support and Behavior Management.

Table 4
Latent Regression Analysis Standardized Coefficients.

	Facilitating Deeper Learning		Supporting Student Expression		Emotional Support and Behavior Management						
	β										
	M1	M2	M1	M2	M1	M2					
Predictors → Process quality											
Total years as KG teacher	-.053	-.097	.041	-.006	-.061	-.076					
Schooling level ^a	.108	.104	.249	***	.247	**	.072	.100			
Training in ECD (yes/no)	.094	.138	†	.172	**	.238	***	.118	†	.117	†
Reading knowledge	.042	.073		.141	†	.166	*	-.008		-.014	
Personal accomplishment	.237	***	.216	**	.137	*	.133	†	.061	.053	
Burnout	-.104	-.095	-.220	**	-.157	*	.054	.065			
Motivation	.202	-.063	-.060		.05		.075	.052			
Job satisfaction	.008	-.014	.153	*	.105		.043	.040			
Classroom resource index	.104	.108	.137	†	.126	†	.260	***	.209	**	
Language used in lesson (ref = local language only)											
English only	.202	.167	-.073		-.005		-.052	.044			
Mixed English and local language	.097	.129	-.034		.048		-.159	-.030			
Teacher is proficient in English	-.041	-.027	-.082		-.079		.122	†	.136	*	
Teacher is proficient in Twi	.055	.018	.028		-.015		.020	.022			
Process Quality → Child outcomes											
Average classroom early literacy skills		.049			.165	**		.075	†		
Average classroom early numeracy skills		.072	†		.072			.076	*		
Average classroom social-emotional skills		.082			.134	*		.018			
Average classroom executive function skills		.038			.072			.067			
Average classroom motor skills		.092			.100	†		.021			

Notes: † $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$. Each column represents a separate regression. Baseline classroom-level scores for all child outcome domains are included in the model, though coefficients are not shown.

^a Continuous variable coded as 1 = primary or JHS, 2 = SHS only, 3 = SHS+ but no college degree; 5 = BA degree; 6 = MA or PhD.

The third factor – Supporting Student Expression (SSE) – involves concepts that include regard for student perspectives (i.e., as considering student ideas), as well as encouraging the development of higher-order thinking skills (i.e., encouraging students to reason and problem-solve, connecting concepts to students’ lives outside of the classroom, language modeling). In various contexts around the world, researchers are concerned that “traditional” approaches to education (i.e., teacher-driven, highly-structured, academically-focused) are being applied inappropriately in early education settings and resulting in the “schoolification” of ECE (Choi, 2006), making ECE simply an extension of ineffective primary schools. This is in contrast to “modernist” approaches to early education (i.e., child-centered, activity- and play-based approaches) that address social skills and the unique needs of very young children (Hirsh-Pasek, 2009). In educational systems that have

traditionally prioritized rote instruction and memorization (Choi, 2006; Osei, 2006), this may be an important dimension of process quality that enables students to deepen their learning. Thus, the quality domain of supporting student expression identified in this study may be particularly important in supporting Ghana’s pre-primary educational system to incorporate more modernist approaches to education.

We also provided evidence of associations between teacher- and classroom-level predictors with each of the three domains of teacher-child interactions. Classroom scores on FDL were associated with teachers’ professional well-being. The SSE factor was associated with more predictors than the other two factors, including training and education, professional well-being, and classroom-level school readiness skills. In light of a growing concern of a teacher “motivation crisis” in Sub-Saharan Africa in

general (Bennell & Akyeampong 2007), and in Ghana specifically (Osei, 2006; Moon 2007), the findings suggest that the resources teachers have available, both material and psychological, have implications for the nature of the interactions children experience in the classroom. Finally, the ESBM factor was significantly associated with classroom physical resources, indicating that teachers may rely on physical resources for implementing positive behavior management (e.g., writing-out class rules on a board). Indeed, elements of structural quality have been shown to promote process quality (e.g., NICHD Study of Early Child Care, 2002). In addition, training in ECD was marginally statistically significantly related to this factor, indicating that teachers with training in early childhood development may implement more developmentally appropriate classroom management techniques. Language of instruction did not significantly predict classroom quality. This issue warrants greater attention in the Ghanaian context, where national language policy mandates that preschool instruction should be in children's mother tongues (United States Agency for International Development, 2009) but there are many challenges to implementation (Ankrah, 2015).

Finally, and perhaps most notable, were the findings on the significant associations between classroom quality as measured by the TIPPS and classroom-levels of children's development of skills during pre-primary school. We found that FDL was not associated with any spring outcome score, including academic outcomes as we might have predicted. Both SSE and ESBM predicted academic outcomes (early literacy and numeracy skills), and SSE predicted classroom-levels of children's social-emotional skills. This pattern of associations is consistent with previous research from the U.S. which found that, compared to children enrolled in programs characterized by low-quality and or medium-quality, high quality instructional support was positively associated with academic and behavioral child outcomes, while quality emotional support was only associated with children's behavioral outcomes at the end of preschool (Mashburn, 2008). The associations with average classroom-level child outcomes range from 0.07 to 0.17, slightly larger than prior rigorous research on classroom quality assessing associations at the child level (see Burchinal, 2017). The findings indicated that the TIPPS measure identified some critical elements of the classroom environment in Ghanaian pre-primary classrooms, and that teachers' ability to support student expression during instruction is significant for improved classroom outcomes. Importantly, the validity of the IDELA assessment of child outcomes has been established in multiple other countries (Halpin et al., 2018) and has been used in multiple studies in Ghana (Wolf & McCoy, 2017; McCoy & Wolf, 2018; Wolf et al., 2018).

Overall, the findings suggest that in Ghana, structural aspects of quality may facilitate the implementation of process quality, and that process quality meaningfully predicts children's academic and behavioral outcomes in preschool. With regard to policy debates concerning program regulation and teacher training needs, these results should be considered in light of the fact that findings are associational and not causal. In addition, many other factors may contribute to classroom quality that were not observed in this study—for example, parent-school relationships, supportive head teachers, and ongoing professional development opportunities for teachers. Our findings certainly do not reflect all sources of variation that have been hypothesized to influence quality, and results should be interpreted with this in mind.

Future research is needed in this area, including test-retest reliability and generalizability studies for the TIPPS and other classroom observation tools in LMIC contexts, to better understand how different time intervals (i.e., time of day, time of year, subject taught, class length) affect the measurement of classroom processes. As this body of work develops, important questions

about high-inference measurement will also need to be considered and addressed. In addition, future studies should investigate SSA teachers' own reflections on classroom-observation tools to ensure that teachers' perspectives and voices are incorporated, and that domains of quality in this particular context are comprehensively considered (e.g., Pastori & Pagani, 2017).

4.1. Limitations and conclusions

This study has limitations that are important to note. First, this study is descriptive in nature and the data collected on teachers are cross-sectional, and thus the relations among predictor variables and classroom quality are not causal. We cannot make conclusions about the extent to which regulating aspects of structural quality (e.g., teacher credentials, classroom resources) might achieve observable and meaningful differences in process quality. This limits the ability to make conclusions about targets for change to improve ECE quality in Ghana. Second, this study includes schools only in the Greater Accra Region of Ghana, and thus only represents urban and peri-urban areas in the country only. As a result, the domains of quality and their associations with teacher and child characteristics cannot be generalized beyond this particular region in Ghana, or to countries outside of Ghana. This is an important limitation, as teachers in rural regions of Ghana face additional unique challenges such as large class sizes, higher student poverty rates, and lower levels of training. Examining how these factors relate to teaching practice, and ECE quality, is an important area for future research. Third, some of the scales used in the analysis had relatively low internal reliability, which may have affected the associations that were estimated. Finally, the data collected on language diversity and language of instruction were limited, precluding a deeper investigation of how this issue may affect teacher-child interactions in this context. Nonetheless, the findings provide a platform from which studies in other contexts and countries can build on.

In a fast-growing field of educational impact evaluations in LMICs (IEG, 2012), and a new directive from the Sustainable Development Goal 4 to strive for the provision of high-quality education for all children, the drive to understand “what works” is being felt by governments, NGOs, donors, and the public alike. The “black box” approach to assessing how programs improve learning outcomes is no longer viewed as sufficient (Yoshikawa et al., 2013). Thus, developing the metrics to assess the classroom environment, and its process quality in particular as a key driver for learning outcomes, is increasingly critical. This study is an important step, yet it is only with systematic data collection and analysis in a diverse array of LMICs that this goal will be addressed on a global scale.

Acknowledgements

We thank the UBS Optimus Foundation and the World Bank Strategic Impact Evaluation Fund (SIEF) for direct support for the data collected that was used in this project, and the NYU Abu Dhabi Research Institute for core support to the Global TIES for Children Center, which provides intramural support to the projects' principal investigators' research activities and directly contributed to the project from which the data comes. We also thank Hirokazu Yoshikawa for his thoughtful feedback on an earlier version of this manuscript.

Appendix A.

Table A1
Descriptive Statistics for Each of the Fifteen TIPPS Items Used.

	Item	Mean	SD	Min	Max
Student Ideas and Interests into Consideration	3	1.27	0.64	1	4
Encourages Reasoning/Problem Solving	4	1.69	0.89	1	4
Instructional Materials	5	3.43	0.73	1	4
Connection to Teaching Objectives	6	2.01	1.03	1	4
Connection to Students Daily Life	7	1.64	0.94	1	4
Specific Feedback	8	2.08	0.91	1	4
Language Modeling	9	1.45	0.72	1	4
Scaffolding	10	2.24	1.07	1	4
Positive Environment	11	3.23	0.70	1	4
Sensitivity and Responsiveness	12	2.78	0.76	1	4
Negative Environment	13	1.55	0.75	1	4
Tone of Voice	14	3.00	0.86	1	4
Behavior Management	17	2.78	0.94	1	4
Consistent Routine	18	2.74	0.77	1	4
Student Engagement	19	3.36	0.64	1	4

Table A2
Results from Exploratory Factor Analysis Three-Factor Model.

	Factor loading		
	Factor 1	Factor 2	Factor 3
Student Ideas and Interests into Consideration	.677		
Encourages Reasoning/Problem Solving	.713		
Instructional Materials			.336
Connection to Teaching Objectives	.614		
Connection to Students Daily Life		.623	
Specific Feedback	.568		
Language Modeling		.510	
Scaffolding	.584		
Positive Environment			.762
Sensitivity and Responsiveness			.688
Negative Environment			-.666
Tone of Voice			.709
Behavior Management			.666
Consistent Routine			.565
Student Engagement			.663

Note: Factor loadings below 0.40 are not shown, with one exception. We show the highest factor loading for the one item that did not load on any factor.

References

- Agbenyega, J. S. (2017). Examining early childhood education system in Ghana: How can Bourdieuan theorization support a transformational approach to pedagogy? In M. Fleer, & B. Oers (Eds.), *International handbook of early childhood education* (pp. 673–690). Netherlands: Springer. http://dx.doi.org/10.1007/978-94-024-0927-7_32
- Agezo, K. C. (2010). Why teachers leave teaching: The case of pre-tertiary institutions in Ghana. *International Journal of Education Reform*, 19(1), 51–52.
- Ankrah, O. T. (2015). *Education experts' perceptions of the Ghanaian language policy and its implementation*. Doctoral dissertation. University of Lapland.
- Araujo, M. C., Carneiro, P. M., Cruz-Aguayo, Y., & Schady, N. (2016). *Teacher quality and learning outcomes in kindergarten*. IDB Working Paper Series, # IDB-WP-665.
- Asare, K., & Nti, S. K. (2014). *Teacher education in Ghana: A contemporary synopsis and matters Arising*. SAGE Open, 1–8.
- Auger, A., Farkas, G., Burchinal, M. R., Duncan, G. J., & Vandell, D. L. (2014). Preschool center care quality effects on academic achievement: An instrumental variables analysis. *Developmental Psychology*, 50(12), 2559.
- Barnett, S. (2004). Retrieved 28 October 2016 from: <http://nieer.org/resources/policybriefs/2.pdf>.
- Bennell, P., & Akyeampong, K. (2007). *Teacher motivation in Sub-Saharan Africa and South Asia*. London: DfID.
- Bidwell, K., & Watine, L. (2014). *Exploring early education problems in peri-urban settings in Africa: Final report*. Newhaven, CT: Innovations for Poverty Action.
- Burchinal, M. (2017). Measuring early care and education quality. *Child Development Perspectives* [E-pub ahead of print]
- Central Intelligence Agency. (2016). *The world factbook*. Ghana: Central Intelligence Agency [Retrieved 22 February 2017 from: <http://www.cia-worldfactbook.us/africa/ghana.html>]
- Chesterfield, R. (1997). *Classroom observation tools in improving educational quality project*. IEQ [Retrieved: http://www.ieq.org/pdf/Class_ObsTool.pdf]
- Choi, S. (2006). *Bite off only as much as you can chew: Gambia's policy of early childhood development*. Policy brief on early childhood development. No 34, Paris: UNESCO.
- Clarke-Stewart, K. A., Vandell, D. L., Burchinal, M., O'Brien, M., & McCartney, K. (2002). Do regulable features of child-care homes affect children's development? *Early Childhood Research Quarterly*, 17(1), 52–86.
- Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis*, 25(2), 119–142.
- Collins, D. (2003). Pretesting survey instruments: An overview of cognitive methods. *Quality of Life Research*, 12(3), 229–238.
- Early, D. M., Maxwell, K. L., Burchinal, M., Alva, S., Bender, R. H., Bryant, D., ... & Henry, G. T. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development*, 78(2), 558–580.
- Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology*, 39(2), 291.
- Ghana Education Service. (2012). *Programme to scale-up quality kindergarten education in Ghana*. Accra: Ghana: Ghana Ministry of Education [Retrieved from <https://issuu.com/sabretom/docs/10.12.12.final.version.of.narrative.op>]
- Ghana Statistical Service. (2012). *2010 population and housing census: Summary report of final results*. Accra, Ghana: Ghana Statistical Service [Retrieved from: <http://www.statsghana.gov.gh/docfiles/2010phc/Census2010-Summary-report-of-final-results.pdf>]
- Gladstone, M., Lancaster, G., Umar, E., Nyirenda, M., Kayira, E., van den Broek, N., & Smyth, R. (2010). The Malawi Developmental Assessment Tool (MDAT): The creation, validation, and reliability of a tool to assess child development in rural African settings. *PLoS Medicine*, 7, e1000273.
- Hair, J. F., Tatham, R. L., Anderson, R. E., & Black, W. (1998). *Multivariate data analysis* (5th ed.). London, UK: Prentice-Hall.
- Halpin, P. H., Wolf, S., Yoshikawa, H. Y., Rojas, N., Kabay, S., Pisani, L., & Dowd, A. J. (2001). Measurement invariance of the international development and early learning assessment (idela) across five countries. *Developmental Psychology* (in press).
- Hamre, B. K., & Pianta, R. C. (2001). Early teacher-child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development*, 72(2), 625–638.
- Hamre, B. K., & Pianta, R. C. (2004). Self-reported depression in nonfamilial caregivers: Prevalence and associations with caregiver behavior in child-care settings. *Early Childhood Research Quarterly*, 19, 297–318.
- Hamre, B. K., Pianta, R. C., Downer, J. T., DeCoster, J., Mashburn, A. J., Jones, S. M., ... & Brackett, M. A. (2013). Teaching through interactions. *The Elementary School Journal*, 113(4), 461–487.
- Hamre, B., Hatfield, B., Pianta, R., & Jamil, F. (2014). Evidence for general and domain-specific elements of teacher-child interactions: Associations with preschool children's development. *Child Development*, 85(3), 1257–1274.
- Hanushek, E. A., & Rivkin, S. G. (2012). The distribution of teacher quality and implications for policy. *Annual Review of Economics*, 4(1), 131–157.
- Hanushek, E. A. (1995). Interpreting recent research on schooling in developing countries. *The World Bank Research Observer*, 10(2), 227–246.
- Harter, S. (1982). The perceived competence scale for children. *Child Development*, 53, 87–97.
- Hirsh-Pasek, K. (2009). *A mandate for playful learning in preschool: Applying the scientific evidence*. Oxford University Press.
- Hoves, C., James, J., & Ritchie, S. (2003). Pathways to effective teaching. *Early Childhood Research Quarterly*, 18, 104–120.
- Hu, B. Y., Fan, X., Gu, C., & Yang, N. (2016). Applicability of the Classroom Assessment Scoring System in Chinese preschools based on psychometric evidence. *Early Education and Development*, 27, 714–734.
- IEG. (2012). *World bank group impact evaluations: Relevance and effectiveness*. Washington, DC: World Bank.
- Janus, M., Brinkman, S., Duku, E., Hertzman, C., Santos, R., Sayers, M., ... & Walsh, C. (2007). *The early development instrument: A population-based measure for communities*. Offord Centre for Child Studies, McMaster University.
- La Paro, K. M. L., Hamre, B. K., Locasale-Crouch, J., Pianta, R. C., Bryant, D., Early, D., et al. (2009). Quality in kindergarten classrooms: Observational evidence for the need to increase children's learning opportunities in early education classrooms. *Early Education and Development*, 20, 657–692.
- Lewis, L. (2013). *Is there a role for the private sector in education? Education for global development*. World Bank [Retrieved 17 February 2017 from: <http://blogs.worldbank.org/education/there-role-private-sector-education>]
- Leyva, D., Weiland, C., Barata, M., Yoshikawa, H., Snow, C., & Treviño, E. (2015). Teacher-child interactions in Chile and their associations with prekindergarten outcomes. *Child Development*, 86(3), 781–799.
- Mashburn, A. J. (2008). Quality of social and physical environments in preschools and children's development of academic, language, and literacy skills. *Applied Developmental Science*, 12(3), 113–127.
- Maslach, C., Jackson, S. E., & Leiter, M. P. (1996). *Maslach burnout inventory manual* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.
- McCoy, D. C., & Wolf, S. (2018). Changes in classroom quality predict Ghanaian preschoolers' gains in academic and social-emotional skills. *Developmental Psychology*.
- McCoy, D. C., Peet, E. D., Ezzati, M., Danaei, G., Black, M. M., Sudfeld, C. R., ... & Fink, G. (2016). Early childhood developmental status in low- and middle-income countries: National, regional, and global prevalence estimates using predictive modeling. *PLoS Medicine*, 13, e1002034.
- Millsap, R. E. (2012). *Statistical approaches to measurement invariance*. New York, NY: Routledge.

- Ministry of Education. (2014). *Ghana 2013 national education assessment: Technical report*. Accra, Ghana: Ministry of Education.
- Moon, B. (2007). School-based teacher development in sub-Saharan Africa: Building a new research agenda. *Curriculum Journal*, 18(3), 355–371.
- NICHD Study of Early Child Care. (2002). Child care structure → process → outcome. *Psychological Science*, 13, 199–206.
- Osei, G. M. (2006). Teachers in Ghana: Issues of training, remuneration and effectiveness. *International Journal of Educational Development*, 26, 38–51.
- Osher, D., Sprague, J., Weissberg, R. P., Axelrod, J., Keenan, S., Kendziora, K., et al. (2007). A comprehensive approach to promoting social, emotional, and academic growth in contemporary schools. In A. Thomas, & J. Grimes (Eds.), *Best practices in school psychology* (Vol. 5) (5th ed., Vol. 5, pp. 1263–1278). Bethesda, MD: NASP.
- Pastori, G., & Pagani, V. (2017). Is validation always valid? Cross-cultural complexities of standard-based instruments migrating out of their context. *European Early Childhood Education Research Journal*, 25(5), 682–697.
- Pakarinen, E., Lerkkanen, M. K., Poikkeus, A. M., Kiuru, N., Siekkinen, M., Rasku-Puttonen, H., & Nurmi, J. E. (2010). A validation of the classroom assessment scoring system in Finnish kindergartens. *Early Education and Development*, 21(1), 95–124.
- Phillips, D., Mekos, D., Scarr, S., McCartney, K., & Abbott-Shim, M. (2000). Within and beyond the classroom door: Assessing quality in child care centers. *Early Childhood Research Quarterly*, 15, 475–496.
- Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38(2), 109–119.
- Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., & Barbarin, O. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions? *Applied Developmental Science*, 9(3), 144.
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system (CLASS) manual: K-3*. Paul H. Brookes Publishing Company.
- Pianta, R. C. (2005). A new elementary school for American children. *SRCD Social Policy Report*, 19(3), 4–5.
- Pianta, R. C. (2011). *Teaching children well: New evidence-based approaches to teacher professional development and training*. Center for American Progress.
- Pisani, L., Borisova, I., & Dowd, A. J. (2015). *International development and early learning assessment technical working paper*. Save the children. [Washington, DC, Available at: <http://resourcecentre.savethechildren.se/library/international-development-and-early-learning-assessment-technical-paper>]
- Quist, H. O. (2003). Transferred and adapted models of secondary education in Ghana: What implications for national development? *International Review of Education*, 49, 411–431.
- Rao, N., Sun, J., Ng, M., Becher, Y., Lee, D., Ip, P., et al. (2014). *Validation, finalization and adoption of the east Asia-Pacific early child development scales*. UNICEF.
- Rivkin, S., Hanushek, E., & Kain, J. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417–458.
- Schaffer, E. C., Nesselrodt, P. S., & Stringfield, S. (1994). The contribution of classroom observation to school effectiveness research. *Advances in School Effectiveness Research and Practice*, 133–150.
- Seidman, E., & Tseng, V. (2011). Changing social settings: a framework for action. In M. Aber, K. Maton, & E. Seidman (Eds.), *Empowering Settings and Voices for Social Change* (pp. 12–37). New York: Oxford University Press.
- Seidman, E., Raza, M., Kim, S., & McCoy, J. M. (2013). *Teacher Instructional Practices and Processes System (V.5) – TIPPS*. New York: NYU: Manual and scoring system.
- Seidman, E., Kim, S., Raza, M., Ishihara, M., & Haplin, P. F. (2018). Assessment of pedagogical practices and processes in low and middle income countries: Findings from secondary school classrooms in Uganda. *Teaching and Teacher Education*, 71, 283–296.
- Slot, P. L., Leseman, P. P., Verhagen, J., & Mulder, H. (2015). Associations between structural quality aspects and process quality in Dutch early childhood education and care settings. *Early Childhood Research Quarterly*, 33, 64–76.
- Squires, J., & Bricker, D. (2009). Ages & stages questionnaires. In (ASQ-3). *A parent-completed child monitoring system* (3rd ed.). Baltimore, MD: Brookes.
- Stallings, J., & Kaskowitz, D. (1974). *Follow through classroom observation evaluation 1972-73 (SRI project URU-7370)*. Stanford, CA: Stanford Research Institute.
- Stigler, J. W., Gallimore, R., & Hiebert, J. (2000). Using video surveys to compare classrooms and teaching across cultures: Examples and lessons from the TIMSS video studies. *Educational Psychologist*, 35(2), 87–100.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Needham Heights, MA: Pearson Education Inc.
- Teachstone Training LLC. (2014). *Teacher-child interactions in early childhood: Research summary*. Charlottesville, VA: Teachstone.
- Teresi, J. A. (2006). Overview of quantitative measurement methods: Equivalence, invariance and differential item functioning in health applications. *Medical Care. Special Issue: Measurement in a Multi-ethnic Society*, 44, 39–549.
- Torrente, C., Aber, J. L., Witteveen, D., Gupta, T., Johnston, B., Shivshanker, A., Annan, J., & Bundervoet, T. (2012). *Baseline Report: Teacher Survey Results.* Report. New York University., http://steinhardt.nyu.edu/scmsAdmin/media/users/eez206/OPEQ/OPEQ_Teacher_Survey_BL_Report_DRC_2012-08-14.pdf.
- Tseng, V., & Seidman, E. (2007). A systems framework for understanding social settings. *American Journal of Community Psychology*, 39(3–4), 217–228.
- Tsoulopas, C. N., Carson, R. L., Matthews, R., Grawitch, M. J., & Barber, L. K. (2010). Exploring the association between teachers' perceived student misbehavior and emotional exhaustion: The importance of teacher efficacy beliefs and emotion regulation. *Educational Psychology*, 30(2), 173–189.
- UNESCO. (2008). Countries on the Move. UNESCO Education for All Global Monitoring Report. Retrieved Feb 17, 2017 from: <http://www.unesco.org/education/gmr2008/press/chapter3.pdf>.
- UNESCO. (2015). Education for All Global Monitoring Report Regional Overview: Sub-Saharan Africa. Paris: UNESCO. Retrieved from: http://en.unesco.org/gem-report/sites/gem-report/files/regional_overview_SSA_en.pdf.
- United Nations. (2015). *Transforming our world: The 2030 agenda for sustainable development*. New York: United Nations [Author]
- United States Agency for International Development. (2009). *National literacy acceleration program (NALAP) baseline assessment*. http://pdf.usaid.gov/pdf_docs/Pnadm581.pdf
- United States Agency for International Development. (2010). *Data collection monitoring report: Standards-based classroom observation protocol for Egypt (SCOPE) in GLO and TILO schools*. <http://www.gem2.org/node/190>. USAID.Egypt
- VSO. (2002). *What makes teachers tick? A policy research report of teacher's motivation in developing countries*. Report. London: VSO. http://www.vso.org.uk/Images/What_Makes_Teachers_Tick_tcm79-21000.pdf
- Vandell, D. (2004). Early child care: The known and the unknown. *Merrill-Palmer Quarterly*, 50(3), 387–414.
- Venäläinen, R. (2008). *What do we know about instructional time use in Mali? Assessing the suitability of classroom observation snapshot instrument for use in developing countries*. Washington, DC: HDNED World Bank.
- von Suchodoletz, A., Fäsche, A., Gunzenhauser, C., & Hamre, B. K. (2014). A typical morning in preschool: Observations of teacher-child interactions in German preschools. *Early Childhood Research Quarterly*, 29(4), 509–519.
- Willis, G. B. (2005). *Cognitive interviewing*. Thousand Oaks, CA: Sage.
- Wolf, S. (2018). Impacts of pre-service training and coaching on kindergarten quality and student learning outcomes in Ghana. *Studies in Educational Evaluation*, 59, 112–123.
- Wolf, S., Torrente, C., Frisoli, P., Weisenhorn, N., Shivshaker, A., Annan, J., & Aber, J. L. (2015). Preliminary impacts of a professional development intervention on teacher wellbeing in the Democratic Republic of the Congo. *Teaching and Teacher Education*, 51, 24–36.
- Wolf, S., Aber, J. L., Torrente, C., McCoy, M., & Rasheed, D. (2015). Cumulative risk and teacher well-being in the Democratic Republic of the Congo. *Comparative Education Review*, 59(4), 717–742.
- Wolf, S., & McCoy, D. C. (2017). Household socioeconomic status and parental investments: direct and indirect relations with school readiness in Ghana. *Child Development*. [E-pub ahead of print].
- Wolf, S., Aber, J. L., & Behrman, J. R. (2018). Experimental impacts of the quality preschool for Ghana interventions on teacher professional well-being, classroom quality, and children's school readiness. *Journal of Research in Educational Effectiveness* (in press).
- Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W. T., & Zaslow, M. J. (2013). *Investing in our future: The evidence base on preschool education*. Ann Arbor, MI: Society for Research in Child Development.
- Yoshikawa, H., Wuermler, A. J., Raikes, A., Kim, S., & Kabay, S. B. (2018). Toward High-Quality Early Childhood Development Programs and Policies at National Scale: Directions for Research in Global Contexts. *SRCD Social Policy Report*, 31(1), 1–36.