POLICY RESEARCH WORKING PAPER 8682

Intervening at Home and Then at School

A Randomized Evaluation of Two Approaches to Improve Early Educational Outcomes in Tonga

Kevin Macdonald
Sally Brinkman
Wendy Jarvie
Myrna Machuca-Sierra
Kris McDonall
Souhila Messaoud-Galusi
Siosiana Tapueluelu
Binh Thanh Vu

Abstract

This paper evaluates and compares two randomized interventions in Tonga, one targeting the home environment of children up to age 5 and one targeting the school environment for first and second grade students. The first intervention supports communities to set up and run playgroups that aim to improve caregiver-child interaction at home and ultimately improve children's readiness for school. Among children of mothers without high school education, being in a treatment community positively affects the school readiness literacy domains and overall score by 0.19 and 0.2 standard deviation, respectively, for girls and the literacy domains for boys by 0.17 standard deviation. The second intervention provides teachers with training, materials, and coaching to improve reading instruction practices and students' reading ability. It increased average early grade reading scores by approximately 0.18 standard deviation per year of exposure. Two cohorts of children were potentially exposed to both interventions, providing an opportunity to compare, for the same population of children, the effects of an early childhood intervention with a school-based intervention using a common measure of learning achievement. The school readiness intervention is found to have positive effects on early grade reading scores only among children in the reading intervention's control group; effect sizes of approximately 0.29 standard deviation were found at the end of second grade for children exposed to one year of the school readiness intervention and at the end of first grade for girls exposed to two years of the intervention.

Intervening at Home and Then at School: A Randomized Evaluation of Two Approaches to Improve Early Educational Outcomes in Tonga[1]

Kevin Macdonald, Sally Brinkman, Wendy Jarvie, Myrna Machuca-Sierra, Kris McDonall, Souhila Messaoud-Galusi, Siosiana Tapueluelu, Binh Thanh Vu

---

1. Introduction

Research identifying low-cost interventions to improve learning outcomes is especially pertinent to developing countries where human capital formation is often constrained by poor education quality and scarce resources to invest in education quality. One way to (loosely) categorize interventions to improve learning outcomes is into those that act through the school environment and those that act through the home environment. Research under the first category generally concentrates on pedagogic interventions including teacher training and learning materials as well as on school management (e.g., a recent review by Kremer et al. 2013), while the second category of research typically focuses on interventions to improve early stimulation or nutrition with parent education delivered through education or health centers, home visits or community groups (as in Berlinski & Schady 2015; Nadeau et al. 2011).

In this paper, we present an evaluation and comparison of two randomized interventions in the Kingdom of Tonga, one targeting the home environment and one targeting the school environment. The first intervention supports community leaders to establish volunteer-facilitated playgroups in which children up to 5 years old and their caregivers are exposed to guided play. The primary objective is to improve early stimulation through better caregiver-child interaction at home in order to improve school readiness. The second intervention provides 1st and 2nd grade teachers with training, lesson plans, materials and coaching on reading instruction. It aims to improve teaching techniques and early grade reading outcomes of students.

Both interventions were implemented as randomized-controlled trials, and measures of school readiness and early grade reading outcomes were collected. First, we find that the interventions had positive effects on the outcomes each intervention targeted. The school readiness intervention is found to have positive effects on school readiness for children with less educated mothers, especially girls. Positive effects on the prevalence of some types of caregiver-child activities at home are also found as well as effects on several indicators of community support for school readiness. The reading instruction intervention is found to have positive effects on early grade reading skills as well as on teachers' teaching techniques. Second, we find that the school-readiness intervention, that targets the home environment, also affects subsequent reading

2

skills for select sub-populations. Exposure to both the school readiness intervention and the reading instruction intervention does not yield any additional effect, and the effects of the school readiness intervention are only found in the reading intervention's control group. While the reading intervention has stronger effects on early grade reading outcomes, it is also more expensive per student; the school readiness intervention has relatively high cost effectiveness for the sub-populations it affects.

Our evaluation contributes to several strands of research. The effects of the school readiness intervention add to the emerging evidence on the effectiveness of playgroups. A closely related intervention in Indonesia targeting children aged 3 and 4 found positive effects on mathematics and language tests (Brinkman et al. 2015; Nakajima et al. 2016); though, it differs from the Tongan intervention by having formally trained local community members to be teachers rather than community volunteers to facilitate the groups. Additionally, the Tongan intervention actively focused on parents and their children participating in the groups, not just the children. Non-randomized evaluations of playgroups in Australia have also found positive effects on measures of school readiness (Gregory et al. 2016, Hancock et al. 2015). To our knowledge, our evaluation is the first randomized evaluation of play-based activities that are "community-led" in the sense that the activities are supported by community leaders and all participants including facilitators are community volunteers.

The school readiness intervention also contributes to research on how community participation can affect educational outcomes. In education research, community participation is typically studied from the perspective of school-based management and accountability (as in Barrera-Osorio et al. 2009; Pradhan et al. 2014; Blimpo, Evans & Lahire 2015). However, in the theory behind playgroups' effect on outcomes, community participation relates less to accountability and more to participatory or peer learning, analogous to how community prenatal and maternal women's groups affect nutrition and health outcomes (e.g.: in Prost et al. 2013). This use of a community-based approach contributes also to the call for more research on different modalities of delivering early childhood development services (Dua et al. 2016; Shonkof et al. 2016).

Our findings on the impact of the reading intervention extend research examining effective

methods of in-service teacher training (for a recent review of evidence see Popova, Evans & Arancibia 2016). It builds on an emerging set of randomized-controlled trials showing the effectiveness of scripted lesson plans and teacher coaching on improving fundamental, early grade reading skills including phonemic awareness and oral reading fluency. These have been implemented in several challenging contexts in developing countries including Uganda, Liberia, Kenya (Piper & Korda 2011; Piper, Zuilkowski & Mugenda 2014; Lucas et al. 2014; Kerwin & Thorton 2015) and, as a remedial programme in Papua New Guinea (Macdonald & Vu 2018).

Finally, a distinct feature of our paper is the potential exposure of two birth cohorts to both interventions. While current evidence suggests that early interventions, including early stimulation, are more effective than later interventions (Nadeau et al. 2011), it is rare to be able to compare the effects of both an early childhood intervention and a school-based intervention on the same population of children. This is an important comparison because education policy makers, despite evidence on the effects of early childhood intervention on educational outcomes, tend only to consider school-based pedagogic interventions as a means to improve learning outcomes. From our experience, this reflects partly the legal mandate of education ministries but also an aversion to a perceived risk: education policy makers are often less familiar with the cost effectiveness of home interventions, especially those delivered outside of preschools, and view the effect on learning outcomes as being abstract or indirect. Our evaluation provides a tangible comparison of an early intervention with a school-based intervention by estimating the effects of both interventions on a common measure of learning achievement and a common population. Our evaluation demonstrates that an intervention targeting the home environment, delivered outside the school system, can be as effective on a per cost basis as a school-based pedagogy intervention but only for some sub-populations.

The paper is structured as follows. The next section presents information on the country context and descriptions of the interventions and research design. The effects of the interventions on their targeted outcomes are then discussed in succession followed by the effect of the school readiness intervention on reading outcomes.

2. Context, interventions and research design

Country context

Both interventions were implemented in the Kingdom of Tonga. The country is a small island state with a population of just over 100,000 inhabitants. While its population is small, the country consists of 176 islands of which 40 are inhabited and spans 718 square kilometers. The kingdom is divided into five main island groups: Tongatapu, Vava'u, 'Eua, Ha'apai and the Niuas. The capital of Tonga, Nuku'alofa, is over 2,000 kilometers from its nearest large market, New Zealand, and over 3,000 kilometers from Australia.

The education system is quite small with approximately 17,000 students in primary school covering ages 6 to 11 and 14,500 secondary students aged 12 to 17. Enrollment rates in primary and secondary school are approximately 90 and 80 percent, respectively. Preschools are private, operated by either churches or communities, and data on preschool participation are limited, though UNESCO UIS reports approximately 2,000 students enrolled.

Despite this level of enrollment at the primary level, students' learning outcomes remain a challenge. For example, assessments of early grade literacy in Tonga in 2009 found that only 30 percent of 3rd grade students were able to read fluently for comprehension (World Bank 2012a). Similar outcomes were found in other Pacific Island countries. In Vanuatu, approximately a quarter of 3rd grade students were able to read fluently for comprehension (World Bank 2012b, 2012c). In Kiribati and Tuvalu, 20 percent of 3rd grade children achieved minimum reading comprehension proficiency (World Bank 2017a, b). In Papua New Guinea, early grade reading assessments conducted in four provinces between 2011 and 2013 found that students lagged two years behind curriculum targets for fundamental pre-reading skills (World Bank 2014a, 2014b, 2014c, 2014d).

This poor performance in reading outcomes in Pacific Island countries is the primary motivation

for piloting the two interventions evaluated in this paper. The pedagogic intervention to improve reading instruction responds directly to the deficiencies found in the early grade reading assessments implemented in Tonga and other Pacific Island countries. The school readiness intervention emerged from a number of factors including substantial within-class variation in 1st grade reading performance, low levels of preschool participation, and few financial resources to expand preschool participation sustainably. The Tongan context is shared by other Pacific Island countries but also developing countries more generally, especially with remote populations.

School readiness intervention

The school readiness intervention aims to affect young children's early stimulation at home by improving child-caregiver interactions. The intervention is motivated by the importance of a young child's physical, socio-emotional and cognitive development prior to school for his or her ability to learn in a school environment (Black et al. 2016; Nores & Barnett 2010). The intention of the intervention is to establish playgroups, termed community play-based activities (CPBAs), which consist of children up to age 5 and their caregivers. The CPBAs are led by volunteer facilitators who receive training and mentoring as part of the intervention. They occurred once or twice per week and lasted approximately two hours per session.

The intervention did not establish CPBAs directly; instead, it supported community leaders to initiate and maintain them. To help initiate CPBAs, the intervention team provided information and CPBA start-up materials to community leaders but also worked closely with them to engage community members to generate interest and support for the CPBA. This involved establishing a community education committee, finding facilitators, finding a venue for the activities which were normally town halls, churches or schools, and raising awareness of the CPBA among parents. The intervention team also worked closely with community leaders throughout the intervention period, helping to resolve problems including the loss of venue or facilitator. Ultimately the establishment of a CPBA in a treatment community relied on the community's leader and members. In some treatment communities, CPBAs were not established or were established later during the intervention.

For this evaluation, the intervention is defined as this support provided by the intervention team which was randomized at the community level. The evaluation is at a community aggregate level rather than at the level of an individual child who attended a CPBA. It is not possible to identify the effect at the child level because exposure to CPBA is not random. Communities self-select by deciding whether to establish a CPBA, and parents self-select by choosing to send their children to a CPBA if it is established. For this reason, the evaluation measures the intent-to-treat effect and underestimates the effect of CPBA exposure.

The intervention was expected to affect school readiness and subsequent learning outcomes in school through two stages. First, the intervention aimed to induce communities to establish and support a CPBA. Second, by participating in a CPBA, children were exposed to new learning opportunities and socialization, and caregivers were exposed to new types of play-based activities and interactions with their children that they could repeat at home, thus increasing stimulation for young children in the home environment.

The motivation for the intervention stems from several sources. First, previous research has established a link between early stimulation at home and child development, future educational outcomes and even labor market outcomes (e.g.: Gertler et al. 2014). Second, community group learning has been shown to be effective at affecting parenting behaviors, both in the context of playgroups (Brinkman et al. 2015; Nakajima et al. 2016) but also in nutrition and health as in community mothers' groups for prenatal health and child nutrition (Prost et al. 2013; O'Rourke, Howard-Grabman, & Seoane 1998). The effectiveness of this approach to parent education reflects the benefits of participatory or peer learning and its ability to influence behaviors. Third, communities in Tonga are tight-knit and exhibit substantial social-capital (Farran 2009; Toganivalu 2008; Huffer 2006; Griffen 2006; World Bank 2013). This was expected to not only improve the effectiveness of peer learning approaches but also the ability of the communities to implement and sustain CPBAs.

The intervention began in mid-2015 and completed at the end of 2017; by June 2017, 1,337 children were participating. Because all children aged 0 to 5 in treatment communities are potential beneficiaries of the intervention, the recurrent cost per child per year was US$ 12.62.

This includes the cost of supporting community leaders, training and mentoring but excludes the time of volunteers and any support provided to the CPBA by community members.

Reading instruction intervention

The reading instruction intervention, entitled "Come Let's Read and Write" (CLRW), provided teachers in treatment schools with training and materials to implement a new pedagogic approach for reading instruction. The training and materials were coupled with monitoring and coaching, and offered to 1st and 2nd grade teachers. The approach aligns key learning competencies for basic reading and writing stipulated in the official Tongan curriculum with a greater degree of clarity on the sequence in which these skills should be taught.

Like the school readiness intervention, CLRW was expected to affect student reading outcomes in two stages. First, the intervention provided teachers with training, materials and coaching in order to induce them to adopt the new pedagogic approach. Second, students exposed to the new pedagogic approach were expected to have acquired better reading skills. For this paper, we refer to the former as the CLRW intervention and the latter as the CLRW pedagogic approach.

Evidence that the pedagogic approach can improve students' reading ability stems from research on what skills children need in order to learn alphabetic languages (Wolf 2007; Linan-Thompson and Vaughn 2007; Sprenger and Charolles 2004; Chiappe & Wade-Woolley 2002; Gove & Cvelich 2011 and National Reading Panel 2000). These include, among others, an understanding of the relationship between printed letters and sounds (Scarborough 2002) and the speed at which a child can read (Fuchs et al. 2001; Abadzi 2006). CLRW is designed around the sequencing of skills and pedagogic methods based on this research (e.g.: August & Shanahan 2006; National Institute for Child Health and Human Development 2000; Pressley 1998; Snow, Burns & Griffin 1998).

In order to induce teachers to implement this structured pedagogic approach, the CLRW intervention relied on highly scripted lesson plans and periodic monitoring and coaching. Previous evaluations of similar interventions have shown positive effects on reading skills (Piper

& Korda 2011; Piper, Zuilkowski & Mugenda 2014; Lucas et al. 2014), and one study has shown a desired effect on teaching practices, measured by classroom observations (Kerwin & Thorton 2015). Both the scripted lesson plans and periodic coaching are seen as crucial to the effectiveness of the CLRW intervention. In-service training interventions in developing countries have generally yielded little evidence of impact, and one reason is a focus on knowledge rather than teaching techniques (Popova, Evans & Arancibia 2016).

The intervention was implemented in 2015 in first grade and extended to second grade in 2017 and 2018. 1st and 2nd grade teachers in 38 schools received training and materials, and the annual recurrent cost of the intervention was US$62.57 per student. This cost includes training, materials, monitoring and coaching, but it does not include the cost of developing the pedagogic approach, as this cost would not be incurred again if the intervention was scaled up.

Research design

Both interventions were implemented as cluster randomized-controlled trials. For the school readiness intervention, 45 community clusters consisting of 59 *de facto* communities and 66 statutory villages were randomly selected to receive the treatment while 45 other community clusters consisting of 83 statutory villages were randomly selected as control group communities. In a 2014 school readiness assessment, villages in Tonga were grouped into 129 communities and average school readiness scores were reported for each of these communities. For the intervention, 90 communities were randomly drawn from these 129. The communities were stratified by island group, and within island group, communities were implicitly stratified by the community's number of children and the 2014 school readiness score. These 90 communities were then assigned to either the treatment or control group using systematic sampling, with communities ordered by their 2014 TEHCI score. This approach ensured that the 90 communities drawn to participate in the study had a similar number of children and school readiness score as those not selected, and that the treatment and control communities had similar school readiness scores at baseline.

For the CLRW intervention, 37 primary schools were randomly selected into the treatment group

and 36 primary schools were randomly selected into the control group. Selection was stratified by island group, school ownership (public or private), whether the school received students from communities included in the school readiness intervention, and the number of children in the first 6 grades. One school dropped out of the intervention after the second year, and this was replaced by a randomly selected school. Not all schools were eligible to be selected into the control or treatment groups: two other primary school interventions were being piloted in several schools which are excluded from the evaluation's population. This included all schools on the island of 'Eua. Note that schools in the treatment and control groups could include students from the school readiness intervention's treatment communities, control communities and unassigned communities.

Data were collected on both child outcomes and behaviors that the interventions were trying to influence for both interventions. For the school readiness intervention, the Tongan Early Human Capabilities Index (TEHCI) survey measured school readiness and collected data on home activities, and surveys of treatment and control communities were conducted to measure community activities that supported school readiness. For the CLRW intervention, classroom observations were conducted to measure teaching practices and the Tongan Early Grade Reading Assessment (TEGRA) measured students' early grade reading skill. These data are discussed in more detail in the sections on the respective interventions. Table 1 presents a timeline for the interventions and data collection by birth cohort. These cohorts are approximate, as children may enter first grade later and sometimes earlier than age 6.

As noted above for both interventions, there is a distinction between intervention and treatment. For the school readiness intervention, we are measuring the effect of providing community leaders with support to establish and maintain CPBAs, not the effect of a child's exposure to a CPBA. For the reading intervention, we are measuring the effect of providing teachers with training, materials and coaching, not the effect of exposure to the new pedagogic approach. In both cases, the effect size of the intervention underestimates the effect size of the underlying treatment. As a result, the effects of the two interventions are not directly comparable primarily because all children in the CLRW treatment schools are exposed to the underlying treatment while a much smaller fraction of children in the community are exposed to a CPBA. However,

in the latter case, children are exposed to effects which arise outside the CPBA resulting from increased community support and awareness for school readiness.

Table 1. Cohorts, intervention exposure, and surveys (excludes CPBA admin. and community data)

Year	2014	2015	2016	2017
Interventions		School readiness CLRW grades 1	School readiness CLRW grades 1&2	School readiness CLRW grades 1&2
Surveys	TEHCI Mar TEGRA Nov		TEGRA Feb & Oct	TEHCI TEGRA Mar & Oct
Cohort born 2014 Intervention Data		Age 0 School readiness	Age 1 School readiness	Age 2 School readiness TECHI
Cohort born 2013 Intervention Data	Age 0	Age 1 School readiness	Age 2 School readiness	Age 3 School readiness TEHCI
Cohort born 2012 Intervention Data	Age 1	Age 2 School readiness	Age 3 School readiness	Age 4 School readiness TEHCI
Cohort born 2011 Intervention Data	Age 2	Age 3 School readiness	Age 4 School readiness	Age 5, Grade 1 CLRW TEGRA
Cohort born 2010 Intervention Data	Age 3 TEHCI	Age 4 School readiness	Age 5, Grade 1 CLRW TEGRA	Age 6, Grade 2 CLRW TEGRA
Cohort born 2009 Intervention Data	Age 4 TEHCI	Age 5, Grade 1 CLRW	Age 6, Grade 2 CLRW TEGRA	Age 7, Grade 3 TEGRA Oct.
Cohort born 2008 Intervention Data	Age 5, Grade 1 TEHCI / TEGRA	Age 6, Grade 2	Age 7, Grade 3	Age 8, Grade 4
Cohort born 2007 Intervention Data	Age 6, Grade 2 TEGRA	Age 7, Grade 3	Age 8, Grade 4	Age 9, Grade 5

3. Effect of the school readiness intervention

The school readiness intervention was expected to affect children's school readiness primarily by improving caregiver-child interaction including through early stimulation in the household. The intervention supported community leaders to establish a CPBA, and if a CPBA was established, it provided training and mentoring to the volunteer facilitators. By participating in CPBAs, caregivers learn new types of interactions that they can do at home with their children that promote school readiness. Children participating in the CPBAs are also exposed to socialization and new learning opportunities. This section presents the effects of the school readiness

intervention. We find that the intervention positively affects school readiness outcomes for children with mothers who are less educated, especially girls. The intervention increases the prevalence of some home activities as well as measures of community support for school readiness.

Data

Two sources of data are used to evaluate the school readiness intervention. First, community monitoring data were collected in treatment and control communities about their activities and support of early childhood care and education (ECCE). Information was collected about whether the community had meetings about ECCE, whether there was a community education committee, whether the committee had met, what type of support it received, whether there were playgroups or a preschool, and health services and activities, among others. Respondents included district officers, town officers, education committee members if one existed, and parents. However, because the purpose of the questionnaire was to monitor community support for playgroups in treatment communities and for ECCE more broadly in control communities, the questionnaires administered to treatment and control communities differ. Eight variables were identified as being relevant to the objectives of the intervention and compatible between the two questionnaires, these are the following: whether a preschool exists in the community, whether there is a play group including a CPBA, whether communities support ECCE services by encouraging parents, fundraising or by providing material resources, whether the community has an education committee, whether the community has a health center where the district nurse is located or centered, and whether the community has an education committee which includes the district nurse as a member.

The second set of data is the Tonga Early Human Capabilities Instrument (TEHCI) survey that measures school readiness for children aged 3 to 5 across a range of developmental domains. The survey was developed in 2013 by defining school readiness based both on the international evidence of the cognitive, socio-emotional and physical development needed to succeed in a school environment but also on Tongan values identified through consultations and workshops. The instrument measures development across eight developmental domains: physical

12

development, verbal skills, culture and spirituality, social and emotional development, perseverance, approaches to learning, numbers and concepts, and literacy skills. It is a rating type assessment where information about a child's abilities and behaviors is obtained from individuals (including caregivers or teachers) that know the child well. The instrument is not designed to provide a diagnostic of a child's development but rather to make inferences about developmental outcomes of a population.

The survey also collected data on the frequency of children being exposed to six different types of activities at home. These variables provide some insight into the home environment, though they do not provide information about the duration or quality of these caregiver-child interactions. The average was used as an index of home activities for the analysis.

The survey was conducted in 2014 prior to the interventions and in 2017 near the end of the interventions. In both rounds, all communities in Tonga were sampled and all children aged 3 to 4 who were identified by local informants including district nurses were sampled. All 5-year-olds were also sampled in 2014, but only those who were not in primary school were sampled in 2017. In this latter sample, the number of 5-year-olds is approximately 40 percent of the number of 3- or 4-year-olds. As a result, the 3- and 4-year-old sample is used to estimate the impact of the community school readiness interventions on school readiness outcomes and support for school readiness at children's homes. Figure 1 presents a density plot of the average TEHCI score and the average score for the literacy sub-domains in standard deviations.

For estimates using both types of data, data are weighted by the inverse of communities' selection probabilities, and standard errors are adjusted using a finite population correction to account for the small number of communities in Tonga from which the treatment and control communities were randomly drawn. Because the TEHCI survey samples all children in each community, there is no sampling variation within communities.

Figure 1. Density plots of average TEHCI score and literacy score (in standard deviations)
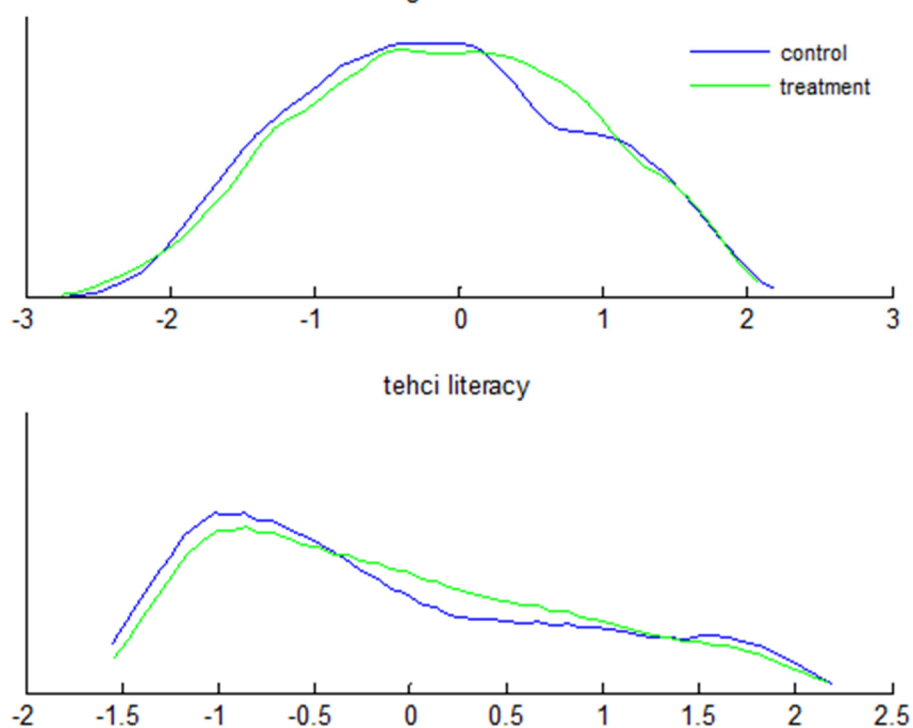
Baseline balance

Table 2 presents comparisons between control and treatment communities in key indicators of the 2014 TEHCI for children aged 3 to 5. No statistically significant differences are found between treatment and control communities except for the proportion of girls, which is 3 percentage points higher in treatment communities. The effects of the school readiness intervention are estimated using the 2017 round of TEHCI. The two surveys were not implemented as panels, as children aged 3 to 5 in the 2017 round would not have been sampled in the 2014 round. No baseline data are available for community monitoring data.

Table 2: TEHCI descriptive statistics and baseline balance (community level)

	all	treatment	control	difference
average TEHCI score	0.051	0.043	0.059	-0.016
	(0.024)	(0.038)	(0.031)	(0.049)
average literacy domains (SD)	0.063	0.046	0.08	-0.035
	(0.02)	(0.032)	(0.024)	(0.041)
home activities index	0.785	0.781	0.789	-0.008
	(0.007)	(0.01)	(0.009)	(0.013)
attending a preschool	0.415	0.401	0.43	-0.029
	(0.012)	(0.018)	(0.017)	(0.026)
female	0.464	0.479	0.449	0.03***
	(0.005)	(0.007)	(0.008)	(0.01)
mother with high school edu.	0.596	0.579	0.615	-0.036
	(0.012)	(0.018)	(0.017)	(0.025)

Standard errors included in parentheses. For differences, statistical significance at the 1, 5 and 10 percent levels are denoted by ***, **, and *, respectively.

Empirical strategy

The randomized assignment of communities to receive the intervention allows us to identify the effect of the intervention; self-selection would require moving between communities which was unlikely. The effect of the intervention on school readiness was expected to arise through family participation in the CPBAs. We cannot estimate the effect of the CPBAs because communities self-select whether they establish a CPBA and families self-select participating in CPBAs. One approach would have been to use the randomized assignment of the intervention as an instrument for estimating the effect of the CPBAs; however, we believed the exclusion restriction would be violated. The establishment of the CPBA involved creating education committees and raising awareness in communities which could affect parenting for families that do not participate in CPBAs. As such, we have analyzed the data as intent to treat at the community level,

In our model, the community intervention, T_{Comm}, provided communities with support and

15

awareness to establish a community-based play activity, and it was expected to impact three sets of outcomes: (1) community factors including discussion of school readiness at community meetings and the establishment of CPBAs, (2) participation in CPBAs and caregiver-child activities at home, and (3) children's school readiness. The effects of the school readiness intervention on each of these outcomes, Y_{SR}, are estimated by

$$Y_{SR} = \beta_0^1 + \beta_1^1 T_{Comm} + u_1 \tag{1}$$

We are interested in how the school readiness intervention affects children from disadvantaged socio-economic groups. One indicator of this is mother's education. The report of the 2014 TEHCI found that children of mothers without high-school education had lower school readiness measures (Brinkman & Vu: 31). More broadly, mother's socioeconomic status has been linked with better educational outcomes (Bjorklund, Lindahl & Plug 2006), health and child development including obesity (Currie 2009), health for young adults (Lundborg, Nilsson & Rooth 2014), age of marriage and fertility (Breierova & Duflo 2004), among others. For a child's participation in a CPBA, caregiver-child interaction, and school readiness, the effects of the school readiness intervention are estimated for different sub-populations defined by gender and whether the child had a mother with a high-school education or not. Children of mothers without a high-school education comprise 30 percent of our sample. The effects by sub-population are estimated with the following model:

$$Y_{SR} = \beta_0^2 + \beta_1^2 T_{Comm} + \beta_2^2 f + \beta_3^2 f T_{Comm} + \beta_4^2 h + \beta_5^2 h T_{Comm} + \beta_6^2 f h T_{Comm} + u_2 \tag{2}$$

where f is a binary variable for being a girl, and h is a binary variable for having a mother who completed high-school.

Effects of the school readiness intervention

Table 3 presents the effect of the school readiness intervention on the measures of community support for school readiness collected in mid-2017. These are estimated with model (1). Data were collected previously, but this table is shown because it is the last round of data collected

and best reflects the intervention when fully implemented. Effects on the previous rounds of data collection are larger, as control communities had lower measures of support. Positive effects are found for whether the community had a playgroup or CPBA, whether the community supported ECCE through fundraising or donating materials, whether the community had an education committee, and whether the community had set up a health center for the district nurse. Treatment communities were slightly less likely to have an education committee that included the district nurse. Overall, the intervention had a positive effect on these measures of support for school readiness by the community. Note that these are not measures of implementation of the intervention, as the intervention does not guarantee change in these measures.

Table 3: Effect of the community intervention on community support for school readiness

	treatment	control	difference
Whether a preschool exists in the village	0.3	0.26	0.039
	(0.034)	(0.032)	(0.047)
Whether there is a playgroup in the village (including CPBA)	0.804	0.017	0.787***
	(0.029)	(0.009)	(0.03)
Community supports ECCE services by encouraging parents	0.832	0.835	-0.003
	(0.024)	(0.019)	(0.039)
Community supports ECCE services by fund raising	0.303	0.188	0.115**
	(0.033)	(0.024)	(0.047)
Community supports ECCE services by providing resources	0.217	0.135	0.082*
	(0.03)	(0.02)	(0.041)
Community has an education committee	0.896	0.559	0.337***
	(0.022)	(0.033)	(0.041)
Community has a health centre where the district nurse is located	0.207	0.137	0.07*
	(0.03)	(0.025)	(0.039)
Community has an educ. cmtte. and the district nurse is a member	0	0.017	-0.017*
	(0)	(0.009)	(0.009)

Standard errors included in parentheses. For differences, statistical significance at the 1, 5 and 10 percent levels are denoted by ***, **, and *, respectively.

The effects of the school readiness intervention on the outcomes measured in TEHCI are presented in Table 4. The effects of the intervention on outcomes for all students are estimated using equation (1) and for the sub-populations of students using equation (2) by least squares. Full model estimates are presented in Annex Tables 1 and 2. Positive effects on overall school readiness were found only for girls of mothers with no high-school education; positive effects are found for both boys and girls of mothers with no high-school education on the average literacy domains. Annex Table 3 presents the effects for each TEHCI school readiness domain. For all children, positive effects were found only on the cultural and spiritual domain and the numeracy concepts domain. For girls of mothers without high-school education, the intervention also affected socio-emotional development, approaches to learning, and writing. Some negative effects were found for boys of mothers with high-school education.

Table 4. Effect of the school readiness intervention

	all	mothers without high school education		mothers with high school education	
		girls	boys	girls	boys
average TEHCI score (SD)	0.04	0.2**	0.08	-0.02	-0.07
	(0.06)	(0.1)	(0.08)	(0.07)	(0.05)
average TEHCI literacy domains (SD)	0.07	0.19**	0.17**	-0.02	-0.03
	(0.05)	(0.08)	(0.07)	(0.06)	(0.05)
home activities index	0.03	0.01	0.05	0.01	0.02
	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)
attended a CPBA	0.21***	0.24***	0.2***	0.21***	0.19***
	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)
attended preschool	-0.02	0.02	0.06**	-0.07**	-0.05**
	(0.02)	(0.04)	(0.03)	(0.03)	(0.03)

Standard errors included in parentheses. Statistical significance at the 1, 5 and 10 percent levels are denoted by ***, **, and *, respectively. Effects are estimated by equations (1) and (2), presented in Annex Tables 1 and 2. Effects are shown for children aged 3 and 4. Sample sizes are 3,429 children except for the home activities index model estimation which is 3,426 children.

No effects were found on the index of home activities; however, as shown in Annex Table 4, positive effects were found for individual home activities including reading to the child, singing songs, and naming or counting things. Positive effects were found for whether a child attended a

CPBA; this is nearly equivalent to the proportion of children attending a CPBA in treatment communities, as very few children in control communities attended a CPBA or playgroup. Approximately 21 percent of 3- and 4-year-olds in treatment communities attended CPBAs. Finally, the school readiness intervention had a positive effect on preschool participation for boys of mothers without high-school education but had a negative effect on those of mothers with high-school education.

Cost effectiveness

The annual recurrent cost of the intervention was US$ 67,802.75. This included visits by the intervention team to support community leaders in setting up and running CPBAs as well as training and mentoring of volunteer facilitators. This cost excludes the opportunity cost of the facilitators' time as well as any donations, either in-kind or monetary, provided by community members. Effect sizes of the school readiness intervention are per child, regardless of whether they attended a CPBA. All children in treatment communities aged 0 to 5 are potential beneficiaries. The number of children aged 0 to 5 in treatment communities was estimated to be 5,373 based on the 2017 TEHCI. The subsequent annual recurrent cost per child was US$ 12.62.

A common measure of cost effectiveness in the impact evaluation literature is standard deviations per US$100 (e.g.: Abdul Latif Poverty Action Lab 2014). Table 5 presents the effect sizes of the intervention in terms of a standard deviation per US$100 per child; this is equivalent to the effects presented in Table 4 divided by the cost per child in 100s of US$. 3- and 4-year-olds sampled in the 2017 TEHCI would have been exposed to the school readiness intervention for approximately two years. The cost per child of two years' exposure was US$ 25.24. Positive effects were found on overall school readiness among girls of mothers with no high-school education and on the average of the literacy domains for both girls and boys of mothers with no high-school education. The effect size per 100 US$ for these sub-populations was 0.79, 0.75 and 0.67 standard deviations, respectively.

Discussion

The school readiness intervention was found to have a positive effect on overall school readiness measured by the TEHCI for girls of mothers with no high-school education. It also affected the school readiness literacy domains of girls and boys in this group. Children of mothers without high-school education had lower levels of school readiness compared to children of higher educated mothers in the baseline TEHCI; hence, the school readiness intervention seems to have benefited the more vulnerable children.

Table 5. Effect size per 100 USD

	all	mothers without high school education		mothers with high school education	
		girls	boys	girls	boys
average TEHCI score	0.16	0.79**	0.32	-0.08	-0.28
	(0.48)	(0.79)	(0.63)	(0.55)	(0.4)
average TEHCI literacy domains	0.28	0.75**	0.67**	-0.08	-0.12
	(0.4)	(0.63)	(0.55)	(0.48)	(0.4)

Standard errors included in parentheses. Statistical significance at the 1, 5 and 10 percent levels are denoted by ***, **, and *, respectively. Estimates are calculated by dividing the effects presented in Table 4 by the per child cost of the intervention in 100s.

No effects were found for the index of home activities, though positive effects were found for specific activities. Of course, the data do not provide a comprehensive picture of the quality of the activities or caregiver-child interaction more generally. Effects on school readiness likely reflect changes in the home environment that are not specifically captured in the TEHCI data. The lack of effect on school readiness for mothers with high-school education may be because the home environment is better for this sub-population, mitigating the effect of CPBA participation on the home environment. The six home activities are more prevalent among mothers with high-school education in the baseline, 2014 TEHCI.

While we cannot attribute these effects exclusively to CPBA participation, it is likely that CPBA is the main source of effect. Given the low participation in CPBAs, the effect of participating in a CPBA is likely quite high. For example, the intervention increased overall school readiness of

girls of mothers with no high-school education by 0.2 standard deviation; if this effect were exclusively the result of attending a CPBA, then the effect of CPBA attendance would be 0.83 standard deviation because 24 percent of these children attended a CPBA. Effect sizes may also be understated by the exclusion of 5-year-olds, as 5-year-olds at the time of the TEHCI survey would have had the most exposure to CPBA in 2017 and preceding years. Nevertheless, even with the low participation rate in CPBAs, just the intent-to-treat can have non-trivial effects on children with less educated mothers.

For boys of mothers with high-school education, the intervention reduced participation in preschool and had a small negative effect on some domains of school readiness. It is not clear why these two effects emerge. One hypothesis is that parents may have viewed CPBA as an alternative to preschool, and the reduction in preschool participation reduced school readiness. Preschools generally charge fees for participation while CPBAs did not. For boys of mothers without high-school education, the intervention increased preschool participation.

4. Effect of the CLRW intervention

The CLRW intervention was expected to improve 1st and 2nd grade students' reading skills by changing teaching techniques and exposing students to a new pedagogic approach. The randomized assignment of schools allowed us to identify the effect of the intervention on teaching practices measured in classroom observations and on reading outcomes. Self-selection could occur if students switched schools in response to the intervention; however, no evidence of this was reported. We found that the CLRW intervention had a positive effect on the prevalence of teaching practices promoted by the intervention and a positive effect on students' reading abilities. Effect sizes tended to be higher for those exposed to two years of the intervention rather than just one.

Data

Two sources of data were used to evaluate the effects of the CLRW intervention. The first was based on classroom observations that were conducted approximately halfway through the 2016

and 2017 school years in both CLRW treatment and control schools. Monitors attended reading instruction classes and assessed whether specific teaching practices and techniques promoted by the CLRW intervention were used by the teacher. Practices used by teachers during teaching and observed by the monitors were divided into 6 domains: phonemic awareness, phonics, reading of words, writing words and letters, forming of sentences and reading comprehension. For each of the practices within these domains, the observer rated whether the practice was fully, partially or not implemented. Classroom observations were conducted for all teachers in both 1st and 2nd grade of treatment and control schools. Estimates of effects described below account for unequal selection probabilities of schools and the small population of schools through a finite population correction.

The second source of data, the Tongan Early Grade Reading Assessment (TEGRA) was an adaptation of the Early Grade Reading Assessments initiated by the Research Triangle Institute (RTI International 2009) and adapted and applied to more than 100 languages in 65 countries (Dubeck & Gove 2015). The assessment was adapted to the Tongan language prior to its first implementation in 2009 and updated for subsequent rounds in 2014, 2016 and 2017. The assessment measured reading skills across 9 domains. Data were also collected on several measures of home support for reading.

The 2016 and 2017 rounds of TEGRA were used in this analysis to measure the impact of CLRW on students' acquisition of reading skills during the school year. The survey was conducted in February and October of 2016, corresponding to the beginning and end of the 2016 school year, and in March and October 2017, corresponding to the beginning and end of the 2017 school year. All treatment and control schools were sampled. Within each school, all first and second grade classes were selected within which a random sample of students was drawn, stratified by gender and, for first grade, whether they had attended a CPBA previously. For the October 2017 round, third grade classes were also selected. For this analysis, standard errors were estimated to be robust for intra-cluster correlation within schools (the primary sampling unit) and adjusted for small population size both at the school stratum and student stratum levels. Sampling weights were calculated to account for unequal selection probabilities of schools and attrition of schools and students.

TEGRA provided data on nine reading domains; a list of these is included in the annex. An average reading score is used to measure the overall effect of the CLRW intervention. It was defined as the average of the standardized scores of the nine reading domains. The reading domains were standardized using the mean and standard deviation of the baseline sample for each particular grade and round of survey, as some reading domains were not comparable between rounds. An index of home support was also used in this analysis and defined as the average for the 7 measures of home support. These measures of home support were whether a student receives help with his or her homework, whether parents are interested in his or her school day, whether he or she reads aloud at home, whether someone reads to the student at home, whether the child has been absent in the past two weeks of school, and whether he or she attended preschool and a CPBA. Note that the student is the respondent to these questions and is typically 6- or 7-year-old; it is not clear how reliable their responses were.

Three samples were used to estimate the effects of the CLRW intervention. The first sample was used to measure the average effect of one year of exposure to the intervention. This consisted of 1st grade students in 2016 and 2017. Both cohorts were weighted equally to provide an average effect. 1st grade students in 2015 were excluded from this because no data were collected at the beginning or end of the 2015 school year. The second sample was used to measure the effect of two years of exposure. This consisted of students who started 1st grade in 2016. The third sample was used to measure the persistence of the CLRW intervention's effect on reading skills. This consisted of students who were in 1st grade in 2016 but compared their test scores at the end of the 2016 school year (at the end of two years of exposure to the intervention) with those at the end of the 2017 school year, providing an estimate of the effect of being out of the intervention for one year.

Descriptive statistics and baseline balance

Table 6 presents baseline means and balance of three variables: average TEGRA score, whether a child was female, and the index of home support. Figure 2 plots the density of average TEGRA score and Annex Figure 1 plots the densities for all 9 reading domains. The sample

used to measure the persistence of effect did not have baseline data, as no data were collected for this cohort prior to the intervention. For both interventions, small, statistically significant differences between treatment and control schools were found for average TEGRA achievement and gender; treatment schools tended to have slightly lower baseline achievement and a slightly lower proportion of female students.

Table 6. TEGRA descriptive statistics and baseline balance by sub-sample

	all	treatment	control	difference
1 year exposure sub-sample (students starting 1st grade in 2016 and 2017)				
average TEGRA score (SD)	-0.036	-0.064	0	-0.064**
	(0.013)	(0.018)	(0.017)	(0.025)
female	0.44	0.413	0.475	-0.062***
	(0.004)	(0.003)	(0.006)	(0.007)
home support index	0.528	0.521	0.538	-0.017
	(0.006)	(0.008)	(0.007)	(0.011)
2 year exposure sub-sample (students starting 1st grade in 2016)				
average TEGRA score (SD)	-0.044	-0.072	0	-0.072**
	(0.014)	(0.018)	(0.024)	(0.03)
female	0.407	0.367	0.471	-0.104***
	(0.005)	(0.003)	(0.008)	(0.008)
home support index	0.528	0.521	0.538	-0.017
	(0.006)	(0.008)	(0.007)	(0.011)

Standard errors included in parentheses. For differences, statistical significance at the 1, 5 and 10 percent levels are denoted by ***, **, and *, respectively.

Attrition

Table 7 presents attrition rates between treatment and control schools for samples used to estimate 1 and 2 years of exposure to the CLRW intervention. For the 1-year exposure sample, the attrition is 13 percent overall. Treatment schools have 4.4 percent lower attrition. Attrition rates are higher for the 2-year exposure sample at 33 percent; attrition in treatment schools is 8 percent lower. Comparison of baseline average reading achievement between attritors and non-attritors is presented in Annex Table 5; attritors tend to have lower reading achievement than

non-attritors in for the 1-year exposure sample and no statistically significant difference for the 2-year exposure sample.

Figure 2. Density plots of average TEGRA score by sample and treatment status (in standard deviations)

baseline

endline

Sample 1: 1 year exposure sub-sample
Sample 2: 2 year exposure sub-sample
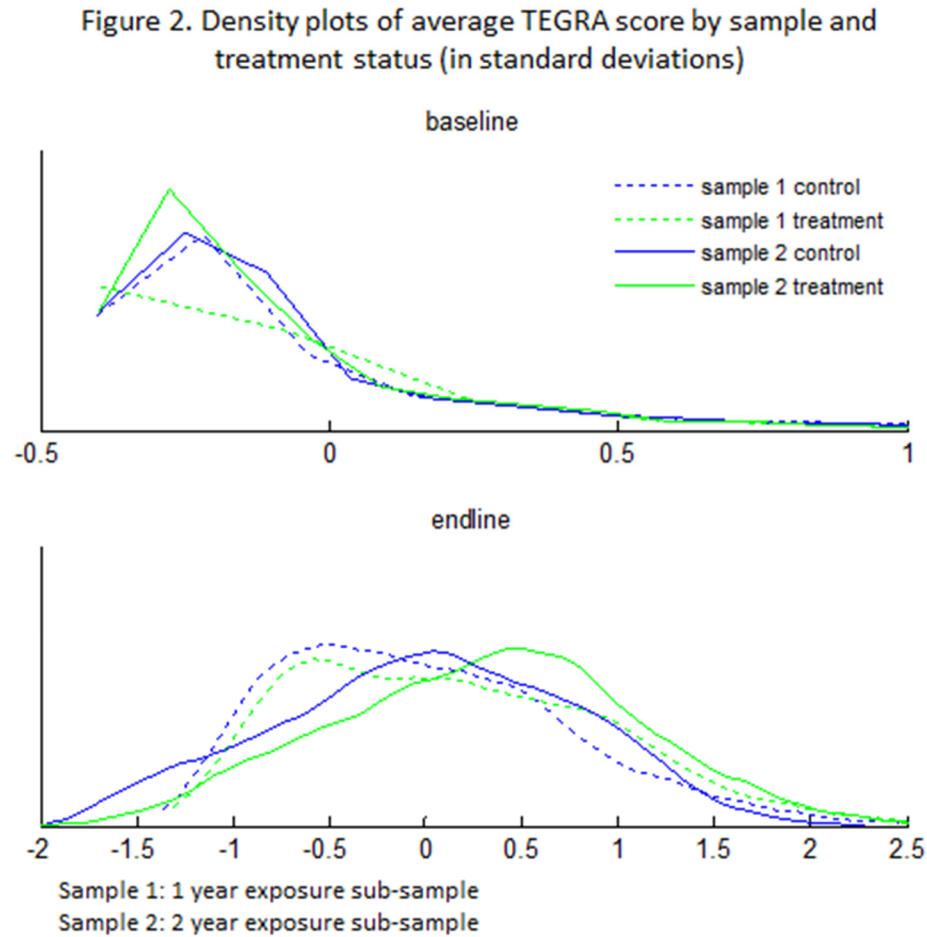
Baseline comparison of the non-attrition sub-samples of the 1- and 2-year exposure samples are also presented in Table 7. The differences between treatment and control groups at baseline for the non-attrition sub-sample are similar to those of the full sample presented in Table 6.

Empirical strategy

The CLRW intervention was expected to affect teaching practices measured in the classroom observations and reading skills measured in TEGRA. The impact of CLRW, T_{CLRW}, on teaching practices, Y_P, is modeled as

$$Y_P = \beta_0^3 + \beta_1^3 \, T_{CLRW} + u_3 \tag{3}$$

Because we were interested in the accumulation of reading achievement during either 1 or 2 years of exposure to the CLRW intervention and because baseline treatment school students tended to have lower achievement than control group students, a value-added model was specified as

$$Y_R^1 = \beta_0^4 + \beta_1^4 \, Y_R^0 + \beta_2^4 \, T_{CLRW} + u_4 \tag{4}$$

where Y_R^1 denotes end-line reading achievement and Y_R^0 denotes baseline achievement.

A difference-in-differences approach was not used because several reading domains were found to be not comparable between baseline and end-line rounds. Finally, we were interested in gender differences in the effect of the CLRW intervention on reading achievement. This was estimated using

$$Y_R^1 = \beta_0^5 + \beta_1^5 \, Y_R^0 + \beta_2^5 \, T_{CLRW} + \beta_3^5 f + \beta_4^5 f \, T_{CLRW} + u_5 \tag{5}$$

Because of differences in attrition between treatment and control groups, estimates of effect may have been biased if the magnitude of effect on a student was correlated with his or her likelihood of attrition. As a result, effects were bounded following Lee (2009). These provide the highest and lowest effect sizes possible given the possibility for bias resulting from the differences in attrition between treatment and control groups. The estimated bounds were based on the value added between baseline and end-line using the estimates from the above equations. For example, the average value-added from baseline to end-line was calculated as $Y_R^1 - \hat{\beta}_1^4 \, Y_R^0$ where $\hat{\beta}_1^4$ is the estimated coefficient for β_1^4. Lee bounds estimates were implemented using a routine by Tauchmann (2012) and modified to accept jackknife replicate weights.

Table 7. TEGRA attrition rates and balance of the non-attrition sample by sub-sample

	all	treatment	control	difference
Attrition rates				
1 year exposure sub-sample				
sample size at baseline	2540	1268	1272	n.a.
sample size at end-line	2199	1226	1073	n.a.
attrition rate (pop. est)	0.13	0.111	0.155	-0.044***
	(0.006)	(0.008)	(0.008)	(0.012)
2 year exposure sub-sample				
sample size at baseline	1239	620	619	n.a.
sample size at end-line	818	383	184	n.a.
attrition rate (pop. est)	0.331	0.299	0.381	-0.081***
	(0.013)	(0.017)	(0.018)	(0.025)
Non-attrition sample baseline balance				
1 year exposure sub-sample				
average TEGRA score (SD)	-0.026	-0.059	0.018	-0.078***
	(0.014)	(0.019)	(0.019)	(0.027)
female	0.44	0.414	0.475	-0.061***
	(0.004)	(0.004)	(0.007)	(0.008)
home support index	0.526	0.52	0.538	-0.018
	(0.006)	(0.009)	(0.008)	(0.012)
2 year exposure sub-sample				
average TEGRA score (SD)	-0.034	-0.064	0.021	-0.085**
	(0.017)	(0.018)	(0.032)	(0.038)
female	0.423	0.392	0.478	-0.086***
	(0.006)	(0.008)	(0.01)	(0.012)
home support index	0.537	0.536	0.539	-0.003
	(0.007)	(0.009)	(0.009)	(0.013)

Standard errors included in parentheses. For differences, statistical significance at the 1, 5 and 10 percent levels are denoted by ***, **, and *, respectively.

Effect of the CLRW intervention

The effect of the CLRW intervention on the percent of teachers fully implementing the teaching practices promoted by the intervention is presented in Table 8. The effects for each category of teaching practice are presented in Annex Table 6. Statistically significant, positive effects were

found both for all teaching practices as well as all sub-domains of teaching. In most domains, effect sizes tended to be higher in the 2016 classroom observations than the 2017 classroom observations. The proportion of teachers in control schools fully implementing the teaching techniques increased between 2016 and 2017.

Table 8. Effect of CLRW on the proportion of teachers fully implementing all teaching practices targeted by the intervention

	Treatment	Control	Difference
School year 2016	0.926	0.518	0.407***
	(0.006)	(0.019)	(0.02)
School year 2017	0.89	0.623	0.267***
	(0.007)	(0.017)	(0.019)

Standard errors included in parentheses. For differences, statistical significance at the 1, 5 and 10 percent levels are denoted by ***, **, and *, respectively.

Table 9. Effect of CLRW on average TEGRA reading scores (standard deviations)

	Effect on students' average TEGRA reading scores		
	1 year exposure	2 years' exposure	1 year after 2 years of exposure
all students	0.19***	0.33***	-0.09**
	(0.04)	(0.06)	(0.04)
girls	0.18***	0.34***	-0.11***
	(0.05)	(0.06)	(0.04)
boys	0.21***	0.38***	-0.04
	(0.04)	(0.08)	(0.05)
difference in girls' and boys' effects	-0.03	-0.04	-0.07
	(0.04)	(0.08)	(0.05)
sample size	2199	818	862

Standard errors included in parentheses. Statistical significance at the 1, 5 and 10 percent levels denoted are by ***, **, and *, respectively. Effects are estimated by equations (4) and (5), presented in Annex Table 7.

Table 9 presents the effects of the CLRW intervention on average TEGRA reading score; these effects were based on the estimates of equations (4) and (5) presented in Annex Table 7. Annex Table 8 presents effects for each domain of TEGRA. Positive effects of 1- and 2-years' exposure to the CLRW intervention were found for all students, and no difference in effects was found between boys and girls. Effect sizes are larger for 2 years of exposure. For the 1-year exposure sample, effects were largest for the initial sounds, letter sounds, reading and listening comprehension domains. Positive effects were not found for the remaining domains. Two years of exposure to the intervention yielded positive effects in all domains except letter names and listening comprehension.

Table 10. Lee bound estimates of impact of CLRW on average TEGRA score

		1 year exposure sample	2 year exposure sample
All students			
	lower bound	0.09**	0.17***
		(0.04)	(0.06)
	upper bound	0.24***	0.51***
		(0.04)	(0.06)
Girls			
	lower bound	0.07	0.15***
		(0.05)	(0.05)
	upper bound	0.24***	0.55***
		(0.05)	(0.07)
Boys			
	lower bound	0.11***	0.2**
		(0.04)	(0.07)
	upper bound	0.25***	0.48***
		(0.03)	(0.07)

Lee (2009) tight bounds are presented. The grouping variables for the 1 year exposure sample are cohort and gender, and for the 2 year exposure sample the grouping variable is gender. Gender is not used as a grouping variable for the bounds estimated for boys and girls. Standard errors included in parentheses. Statistical significance at the 1, 5 and 10 percent levels are denoted by ***, **, and *, respectively. Standard errors are estimated using jackknife replicates.

One year after 2-years of exposure, a negative effect was found for girls but not boys, suggesting that the advantage CLRW provides students may decline for girls; however, the negative effect is smaller than the effect size for either 1 or 2 years of exposure.

Lee (2009) bounds are presented in Table 10 for the effect on average TEGRA achievement. These are "tight" bounds and use gender, and for the 1-year exposure sample, cohort as grouping variables. The student's cohort is also used to tighten the bounds estimated for gender for the 1-year exposure sample. In both cases, the lower bound estimates were positive and statistically significant except for the effect of 1-year of exposure on girls.

Cost effectiveness

The intervention's costs consisted of teacher training workshops, materials and the periodic coaching and mentoring. The annual costs of these totaled US\$87,469.05 for the 37 schools in the intervention. At baseline, 1,398 1st and 2nd grade students attended the treatment schools. The cost per student was US\$ 62.57. Table 11 presents effect sizes of 1 and 2-years of exposure per US\$100. For all cases, the effect sizes per US\$100 are approximately 0.3 standard deviation. While two years of exposure yields higher effect sizes, it is also twice as expensive; hence, the cost effectiveness is slightly less than that of 1 year of exposure.

Table 11. Effect of CLRW TEGRA reading skills per 100 USD

	1 year exposure sample	2 years exposure sample
all	0.3***	0.26***
	(0.06)	(0.05)
girls	0.29***	0.27***
	(0.08)	(0.05)
boys	0.34***	0.3***
	(0.06)	(0.06)

Standard errors included in parentheses. Statistical significance at the 1, 5 and 10 percent levels are denoted by ***, **, and *, respectively. Estimates are calculated by dividing the effects presented in Table 9 by the per student cost of the intervention in 100s.

Discussion

The evaluation finds positive effects both on teachers' teaching practices and average TEGRA achievement. This is consistent with the intervention's expected effect on reading outcomes by first improving the school environment. It is not clear why the percent of teachers fully implementing several of the teaching techniques promoted by the intervention increased in control schools between 2016 and 2017. It may have been a form of contamination in which treatment school teachers were sharing the techniques they had learned with control school teachers; however, control school teachers did not receive the periodic monitoring and coaching that treatment school teachers received, and with the exception of one known case, did not receive the materials. Additional analysis of the data did not find any statistically significant difference in the effect of CLRW on reading outcomes between 2016 and 2017. If there had been some contamination of the control group in 2017, it did not seem to affect the outcomes.

In terms of internal validity, some minor imbalance between treatment and control schools was found in the baseline average TEGRA achievement and gender composition. Our decomposition of effect sizes into gender groups and controlling for baseline average achievement accounts for this imbalance. The validity of our study is also complicated by the high rate of attrition in our samples, especially for the 2-year exposure sample. However, the non-attrition samples did not exhibit a substantially different imbalance than the baseline sample, and the lower Lee bound estimates remained positive and statistically significant except in one case.

The intervention's cost effectiveness was found to be approximately 0.3 standard deviation per US$100. The cost effectiveness of 2 years of exposure was approximately the same as 1 year of exposure.

Finally, the treatment and control school sample's population excluded several schools in Tonga because the government was piloting another intervention. However, it is unlikely that these excluded schools were systematically different. We do not believe that the exclusion of these schools would have had a substantial effect on the external validity of the study.

5. Comparing the interventions' effects on reading skills

The school readiness intervention ultimately aimed to improve the home environment and subsequently improve a child's ability to learn in a school environment. While its objectives were much broader than improving the specific reading skills assessed in TEGRA, an important question is did it have any effect on them? Two birth cohorts were exposed to both the school readiness intervention prior to starting school and the CLRW intervention once in school. Using data on mappings between communities and schools, we estimate the effect of the school readiness intervention on early grade reading scores. We find a positive effect for children with one year of exposure to the school readiness intervention at the end of two years of schooling in CLRW control schools and for girls with two years of exposure to the school readiness intervention at the end of one year of schooling.

Data

In TEGRA, there are two sources of data on whether a student lived in a treatment or control community prior to starting primary school. First, the February round of the 2016 TEGRA collected data on which community 1st grade children lived prior to starting school. However, the information could not be collected for 17 treatment or control schools, and a further review of the data found that it may not have been reliable for all students. Second, 421 students in TEGRA were previously sampled in the 2014 TEHCI data. This source of data of where a child lived prior to primary school was reliable and almost all schools in TEGRA contained some of these students. We use this source to derive for each school the probability that a student was from a school readiness treatment community as opposed to a control community or unassigned community. Unassigned communities were also valid controls as the treatment (and control) communities were selected at random from all communities in Tonga. The indicator was calculated as the proportion of students in a school who came from a CPBA treatment community. This measure was highly correlated with an analogous indicator calculated using the first source of data on students' communities.

A second variable we added was an estimate of the average baseline 2014 school readiness score for students in each school. This was calculated as the average baseline school readiness score for the students in the school that were linked to the 2014 TECHI. This variable was used to establish balance in baseline TEHCI school readiness between treatment and non-treatment students using our mapping approach. An analogous variable was defined using the TEHCI home activity index.

We defined three sub-samples of interest for estimating the effect of the school readiness intervention on TEGRA achievement. The first sub-sample consisted of 1st grade students sampled at the beginning and end of the 2016 school year, and the second sub-sample consisted of 1st grade students sampled at the beginning of the 2016 school year and end of the 2017 school year. Students starting first grade in 2016 would have been exposed to the school readiness intervention in the year prior to starting primary school. This cohort would have been mostly born in 2010. These two sub-samples are used to estimate the effect of one year of exposure to the school readiness intervention on 1 and 2 years of schooling, respectively. The third sub-sample consisted of students in 1st grade during the 2017 school year. They would have been exposed to the school readiness intervention in the two years prior to starting primary school and been born mostly in 2011. This sub-sample was used to estimate the effect of two years of the school readiness intervention and one year of schooling. It should be noted that 2015 was the implementation year of the school readiness intervention, and playgroups did not become established until approximately halfway through the year.

Empirical approach

Let i and j index the ith student in the jth school, and let $T_{SR,ij}$ denote whether the student came from a school readiness treatment community or not. We defined the following model for each school, j, as

$$Y^1_{R,ij} = \beta^6_0 + \beta^6_1 Y^0_{R,ij} + \beta^6_2 T_{SR,ij} + \beta^6_3 T_{CLRW,j} + v_{6,j} + u_{6,ij} \qquad (6)$$

where $u_{6,ij}$ is a mean-zero, student-level error term and $v_{6,j}$ is a school-level error term. If $P_{SR,j}$

is the proportion of students in school j, then averaging (6) by school yields,

$$\bar{Y}_{R,j}^1 = \beta_0^7 + \beta_1^7 \bar{Y}_{R,j}^0 + \beta_2^7 P_{SR,j} + \beta_3^7 T_{CLRW,j} + v_{6,j} \tag{7}$$

where $\bar{Y}_R^1$ is the average of the school's end-line reading achievement, and $\bar{Y}_R^0$ is the average of the school's baseline reading achievement. The estimator of β_3^7 is an unbiased estimate of β_3^6 as long as the proportion of students that are from a school readiness treatment community, $P_{SR,j}$, is uncorrelated with the unobserved school effect, $v_{6,j}$. This is the identifying assumption for this method. It is a reasonable assumption given the randomized assignment of treatment communities. It implies that the school readiness intervention's effect on an individual's reading score is unrelated to the number of students in the school who were exposed to the school readiness intervention.

One approach would have been to estimate (7) using school-level data weighted by the number of students in each school. However, this would have underestimated the sampling error as it ignores within school variation in the estimates of average achievement. Hence, student level data can be used to estimate this equation as

$$Y_R^1 = \beta_0^8 + \beta_1^8 Y_R^0 + \beta_2^8 P_{SR} + \beta_3^8 T_{CLRW} + u_8 \tag{8}$$

We present the effects in terms of their interaction with the CLRW intervention as well as gender using the following equation

$$\begin{aligned} Y_R^1 = \beta_0^9 + \beta_1^9 Y_R^0 + \beta_2^9 P_{SR} + \beta_3^9 T_{CLRW} + \beta_4^9 f + \beta_5^9 f P_{SR} + \beta_6^9 f T_{CLRW} \\ + \beta_7^9 P_{SR} T_{CLRW} + \beta_8^9 f P_{SR} T_{CLRW} + u_9 \end{aligned} \tag{9}$$

Finally, in order to estimate balance between students exposed to the school readiness intervention and those not, the following model is estimated as

$$\bar{Y}_{SR}^0 = \beta_0^{10} + \beta_1^{10} P_{SR} + u_{10} \tag{10}$$

where $\bar{Y}_{SR}^{0}$ denotes estimated average baseline school readiness for the communities mapped to the students in the school. A non-zero estimate of β_2^{11} implies that imbalance exists between the school readiness treatment and control group students using this method of mapping. Equation (10) is also estimated using gender, the TEHCI home activity index, exposure to the CLRW intervention, and attrition as dependent variables, to measure imbalance in these variables as well.

Table 12. Estimates of equation (10) testing balance for the baseline sample

dependent variable:	baseline community TEHCI score (SD)	home support	female	CLRW
1 year of school readiness intervention sample (starting 1st grade in 2016)				
probability of being from a treatment community	-0.06 (0.08)	-0.02 (0.01)	0 (0.01)	0.08 (0.09)
constant	-0.01 (0.04)	0.78*** (0.01)	0.41*** (0.01)	0.58*** (0.05)
observations	1214	1214	1214	1214
r-square	0.00	0.01	0.00	0.00
2 years of school readiness intervention sample (starting 1st grade in 2017)				
probability of being from a treatment community	-0.09 (0.07)	-0.02 (0.01)	0.01 (0.02)	0.11 (0.09)
constant	0 (0.03)	0.78*** (0.01)	0.47*** (0.01)	0.47*** (0.05)
observations	1290	1290	1289	1290
r-square	0.01	0.01	0.00	0.01

Table presents estimates of equation (10) for the listed dependent variables at baseline. Non-zero coefficients of the probability of being from a treatment community imply imbalance between students from treatment and non-treatment communities as a result of the community-school mapping. Standard errors included in parentheses. Statistical significance at the 1, 5 and 10 percent levels are denoted by ***, **, and *, respectively.

Baseline balance

Table 12 presents estimates of equation (10) for the 1-year exposure sample and the 2-year of exposure samples as well as four different dependent variables, in order to test for imbalance between students exposed to the school readiness intervention and those not. Note that the samples used to estimate the effect of one year of exposure to the school readiness intervention on one and two years of schooling are the same at baseline; they started in 1st grade in 2016. No statistically significant association was found between the proportion from a school readiness intervention community and any of the four baseline variables. This suggests no evidence of imbalance.

Attrition

Equation (10) is also estimated using attrition as a dependent variable to test whether there were differences in attrition rates between school readiness treatment and control groups. We find that the school readiness intervention had a negative effect on attrition for only one sub-sample: that used to estimate the effect of one year of exposure to the school readiness intervention on reading scores after one year of schooling. Baseline balance for the non-attrition sample is tested analogously using equation (10) again, and no imbalances were found for any of the sub-samples.

Effect of the school readiness intervention on reading achievement

Table 14 presents the estimated effects of the school readiness intervention disaggregated by gender and CLRW treatment exposure. Model estimates are presented in Annex Table 8. The school readiness intervention had a positive effect on girls' reading achievement after 1 year of exposure to the school readiness intervention and 2 years of schooling as well as after 2 years of exposure to the school readiness intervention and 1 year of schooling, in CLRW control schools only. For this latter group, the effect of CLRW on the effect of the school readiness intervention was negative; being exposed to the CLRW intervention reduced gains from exposure to the school readiness intervention. A positive effect for boys in the CLRW control school after 1

36

year of exposure to the school readiness intervention and 2 years of schooling was also found.

Cost effectiveness

The annual cost per child of the school readiness intervention was US$12.62 per child and for the CLRW intervention, US$62.57 per student. Table 15 provides a comparison of the cost effectiveness of the two interventions. These are based on the effect sizes of exposure to one intervention conditional on no exposure to the other. For example, the effect per US$100 of the school readiness intervention was based on the effect of the school readiness intervention for students in the CLRW control group presented in Table 14, and the effect per US$100 of the CLRW intervention was conditional on there being a zero probability of being from a school readiness intervention community. These figures are based on linear combinations of the estimated coefficients of equation (9) presented in Annex Table 9.

Because the per child cost of the school readiness intervention is much smaller than the per child cost of the CLRW intervention, the effect sizes per US$100 tend to be higher for the school readiness intervention compared to the CLRW intervention. However, the school readiness intervention has positive effects that are statistically significant for three sub-samples, as was presented in Table 14; the CLRW intervention affects reading for all the sub-populations presented here. For one sub-population, girls exposed to one year of the community school readiness intervention and two years of schooling in a CLRW control group, the cost-effectiveness of the school readiness intervention was higher than that of CLRW. For all other sub-populations, no statistically significant difference in cost effectiveness was detected.

Table 13. Estimates of equation (10) testing balance in attrition rates and testing balance for the non-attrition sample

		non-attrition sample			
dependent variable:	attrition	baseline community TEHCI score (SD)	home activity index	female	CLRW
1 year of school readiness intervention 1st grade sample					
probability of being from a treatment community	-0.05**	-0.06	-0.02	0	0.08
	(0.02)	(0.08)	(0.01)	(0.01)	(0.09)
constant	0.14***	-0.01	0.78***	0.41***	0.58***
	(0.01)	(0.04)	(0.01)	(0.01)	(0.05)
observations	1214	1049	1049	1049	1049
r-square	0.00	0.00	0.01	0.00	0.00
1 year of school readiness intervention 1st & 2nd grade sample					
probability of being from a treatment community	-0.02	-0.04	-0.02	-0.01	0.08
	(0.03)	(0.07)	(0.01)	(0.02)	(0.09)
constant	0.25***	-0.01	0.78***	0.41***	0.58***
	(0.02)	(0.04)	(0.01)	(0.01)	(0.05)
observations	1214	925	925	925	925
r-square	0.00	0.00	0.01	0.00	0.00
2 years of school readiness intervention 1st grade sample					
probability of being from a treatment community	-0.02	-0.09	-0.02	0.01	0.11
	(0.02)	(0.07)	(0.01)	(0.02)	(0.09)
constant	0.12***	0	0.78***	0.47***	0.47***
	(0.01)	(0.03)	(0.01)	(0.01)	(0.05)
observations	1290	1140	1140	1140	1140
r-square	0.00	0.01	0.01	0.00	0.01

Table presents estimates of equation (10) for attrition at baseline and for the listed dependent variables at baseline for the non-attrition sample. Non-zero coefficients of the probability of being from a treatment community imply differential attrition rates or, for the non-attrition sample, imbalance between students from treatment and non-treatment communities as a result of the community-school mapping. Standard errors included in parentheses. Statistical significance at the 1, 5 and 10 percent levels are denoted by ***, **, and *, respectively.

Table 14. Effect of the school readiness intervention and CLRW treatment by gender

	1 year of school readiness intervention sample		2 years of school readiness intervention sample
	end of 1st grade average TEGRA score	end of 2nd grade average TEGRA score	end of 1st grade average TEGRA score
girls			
in a CLRW control school	0.17	0.3**	0.29**
	(0.15)	(0.12)	(0.13)
in a CLRW treatment school	0.18	0.09	-0.09
	(0.14)	(0.12)	(0.11)
difference	0.01	-0.21	-0.38**
	(0.21)	(0.17)	(0.17)
boys			
in a CLRW control school	0.12	0.28*	0.01
	(0.13)	(0.15)	(0.1)
in a CLRW treatment school	0.08	0.1	-0.09
	(0.13)	(0.14)	(0.08)
difference	-0.04	-0.17	-0.1
	(0.18)	(0.21)	(0.12)
sample size	1049	925	1140

Standard errors included in parentheses. Statistical significance at the 1, 5 and 10 percent levels are denoted by ***, **, and *, respectively. Effects are estimated by equation (9), presented in Annex Table 9.

Discussion

The school readiness intervention positively affected student reading outcomes for girls in the CLRW control group and, in one case, boys. This reflects the findings in Table 4 showing that girls benefited more broadly than boys from the intervention. The effect size for girls in a CLRW control school exposed to two years of the school readiness intervention and one year of schooling is 0.3 standard deviation, and it is higher than the effect size for girls exposed to one year of the school readiness intervention and one year of schooling. This suggests that the

school readiness intervention enhanced the effect of a year of schooling, and this enhancement is consistent with how the school readiness intervention was expected to affect reading outcomes. Exposure to the school readiness intervention intended to improve the home environment and subsequently a child's preparedness to learn in school.

Table 15. Comparison of effect per US$ 100

	school readiness intervention	CLRW	difference
1 year of school readiness intervention sample			
effect on end of 1st grade reading skills			
females	1.37	0.41***	0.96
	(1.2)	(0.13)	(1.14)
males	0.98	0.44***	0.54
	(1)	(0.12)	(0.95)
effect on end of 2nd grade reading skills			
females	2.36**	0.28***	2.08**
	(0.96)	(0.08)	(0.93)
males	2.19*	0.37***	1.82
	(1.22)	(0.09)	(1.17)
2 years of school readiness intervention sample			
effect on end of 1st grade reading skills			
females	1.17**	0.43***	0.74
	(0.53)	(0.12)	(0.48)
males	0.04	0.3***	-0.26
	(0.38)	(0.1)	(0.35)

Standard errors included in parentheses. Statistical significance at the 1, 5 and 10 percent levels are denoted by ***, **, and *, respectively. Effects are relative to being exposed to neither intervention and calculated by dividing the effects presented in Annex Table 9 by the per student costs in 100s.

No complementarity was found between the school readiness intervention and the CLRW intervention except among girls with two years of exposure to the school readiness intervention; for this sub-population, the effect of the school readiness intervention was found to be smaller in CRLW treatment schools than in control schools. For all other sub-populations, no statistically significant difference in the impact of the school readiness intervention between CLRW

40

treatment and control schools was found. The school readiness intervention was less costly than the CLRW intervention; however, only for one-subpopulation can we conclude that the school readiness intervention was more cost effective.

To estimate the effect of the school readiness intervention on reading scores, we relied on a mapping of students between communities and schools defined at the school level. This provided a probability that a student in a particular school was from a school readiness treatment community. The resulting estimator is unbiased under several conditions. First, if the proportion of students in a school from a treatment community is uncorrelated with the unobserved school effect. This is a reasonable assumption because the assignment of communities was randomized. Second, the probability of a student being from a treatment community is estimated based on a sub-sample of students matched between the TEHCI and TEGRA surveys. Uncertainty of this estimate creates the well-known error-in-regressors bias. If it acts as attenuation bias as in a univariate regression model, then this results in an underestimate of the effect size.

6. Conclusion

This study evaluates two quite different approaches to improve early educational outcomes in Tonga. The school readiness intervention targets the home environment and was implemented by communities; the CLRW intervention targets schools and was implemented more directly by the education ministry. Positive effects for the school readiness intervention were found on children of mothers without a high-school education on the overall school readiness score for girls and the literacy score for both boys and girls of the same mothers. It also affected some specific measures of home activities as well. The CLRW positively affected observed teaching practices in the classroom as well as reading skills for both boys and girls and most domains of reading skill. These findings suggest that expansion of both the CLRW and school readiness interventions is warranted.

Because two cohorts were potentially exposed to both interventions, we are able to compare the effects of these two very different interventions on common measures of learning outcomes for

the same population of children. The school readiness intervention had a positive effect on early grade reading skills but only for select sub-populations and only those in CLRW control schools. No complementarity was found between the two interventions, and for one sub-population, the CLRW intervention reduced the effect of the school readiness intervention. While we find that the school readiness intervention was no less cost effective than the CLRW intervention for improving reading scores, we do not conclude that the school readiness intervention is necessarily an alternative to the CLRW intervention for improving reading outcomes specifically. Rather, comparing these two interventions' effects on the same population of children and outcomes clearly demonstrates the potential of a school readiness intervention, implemented by communities and acting through a complex and indirect chain—via community participation, parental awareness, parenting and the home environment, school readiness and finally learning outcomes—to be among the options for education policy makers to improve learning outcomes more generally.

References

Abadzi, H. 2006. *Efficient learning for the poor: Insights from the Frontier of Cognitive Neuroscience.* Washington, DC: The World Bank.

August, D., and Shanahan, T. 2006. *Developing Literacy in Second-Language Learners: A Report of the National Literacy Panel on Language, Minority Children, and Youth*. Mahwah NJ USA: Lawrence Erlbaum Associates

Barrera-Osorio, F., T. Fasih, and H. A. Patrinos with L. Santibáñez (2009). *Decentralized Decision-Making in Schools: The Theory and Evidence on School-Based Management*. Washington, D.C.: The World Bank.

Berlinski, S. and N. Schady (2015). *The Early Years: Child Well-Being and the Role of Public Policy*. Washington, D.C.: Inter-American Development Bank

Björklund, Anders, Mikael Lindahl, and Erik Plug (2006) "The origins of intergenerational associations: Lessons from Swedish adoption data." *The Quarterly Journal of Economics*: 999-1028.

Black M. M., S. P. Walker, L. C. H. Fernald, C. T. Andersen, A. M. Di Girolamo, C. Lu, D. C. McCoy, G. Fink, Y. R. Shawar, J. Shiffman, A. E. Devercelli, Q. T. Wodon, E. Vargas-Barón, S. Grantham-McGregor, Lancet Early Childhood Development Series Steering Committee (2016). Early childhood development coming of age: science through the life course." *The Lancet*. Jan 7;389(10064):77-90

Blimpo, M. P., D. Evans, N. Lahire (2015). Parental human capital and effective school management : evidence from The Gambia (English). *Policy Research Working Paper* No. 7238 Washington, D.C. : World Bank Group.

Brinkman, S., A. Hasan, H. Jung, A. Kinnell, M. Pradhan. 2015 The Impact of Expanding

Access to Early Childhood Services in Rural Indonesia: Evidence from Two Cohorts of Children. *World Bank Policy Research Working Paper Series*. No. 7372.

Brejerova, Lucia and Esther Duflo (2004). The impact of education on fertility and child mortality: do fathers really matter more than mothers? *NBER working paper series* No. 10513.

Brinkman, Sally and Binh Thanh Vu (2017). *Early childhood development in Tonga: Baseline results from the Tongan Early Human Capability Index*. Washington, D.C.: The World Bank

Chiappe, P., L. Siegel, and L. Wade-Woolley. 2002. Linguistic diversity and the development of reading skills: A longitudinal study. *Scientific Studies of Reading* 6(4): 369–400

Currie, Janet (2009). "Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development." *Journal of Economic Literature* 47, no. 1: 87-122.

Dua, Tarun & Tomlinson, Mark & Tablante, Elizabeth & Britto, Pia & Yousfzai, Aisha & Daelmans, Bernadette & L Darmstadt, Gary. (2016). Global research priorities to accelerate early child development in the sustainable development era. *The Lancet Global Health*. 4. 10.1016/S2214-109X(16)30218-2.

Farran, S. 2009. *Human rights in the South Pacific: Challenges and changes*, 181. London: Routledge Cavendish

Fuchs, L., D. Fuchs, M.K. Hosp, and J. Jenkins. 2001. Oral Reading Fluency as an Indicator of Reading Competence: A Theoretical, Empirical, and Historical Analysis. *Scientific Studies of Reading* 5(3), 239–256.

Gallego, F., E. Näslund-Haldey, and M. Alfonso (2018). Tailorng Instruction to Improve Mathematics Skills in Preschools: A Randomized Evaluation. *IDB Working Paper Series*. No. IDB-WP-905. Washington, D.C.: Inter-American Development Bank

Gertler, P., J. Heckman, R. Pinto, A. Zanolini, C. Vermeerch, S. Walker, S. Chang-Lopez, and S. Grantham-McGregor (2014). Labor Market Returns to an Early Childhood Stimulation Intervention in Jamaica. *Science* 344(6187): 998-1001.

Gregory, T., Harman-Smith, Y., Sincovich, A., Wilson, A., & Brinkman, S. (2016). *It takes a village to raise a child: The influence and impact of playgroups across Australia*. Telethon Kids Institute, South Australia. ISBN 978-0-9876002-4-0.

Griffen V (2006), Gender Relations in Pacific cultures and their impact on the growth and development of children, 'Children's Rights and Culture in the Pacific' Seminar, 30th October 2006

Gove, A. and P. Cvelich. 2011. *Early Reading: Igniting Education for All. A report by the Early Grade Learning Community of Practice. Revised Edition*. Research Triangle Park, NC: Research Triangle Institute.

Hancock KJ, Cunningham NK, Lawrence D, Zarb D, Zubrick SR (2015) Playgroup Participation and Social Support Outcomes for Mothers of Young Children: A Longitudinal Cohort Study. PLoS ONE 10(7): e0133007.

Huffer E (2006), Cultural Rights in the Pacific – What they mean for Children, 'Children's Rights and Culture in the Pacific' Seminar, 30th October 2006

Kremer, M., C. Brannen, and R. Glennerster (2013) The Challenge of Education and Learning in the Developing World. *Science*, Vol. 340 No. 6130 pp. 297-300, April

Lee, D. S. (2009). Training, wages, and sample selection: estimating sharp bounds on treatment effects. *The Review of Economic Studies*. 76(3): 1071-1102

Linan-Thompson, S., and S. Vaughn. 2007. *Research based methods of reading instruction for English language learners: Grades K–4*. Alexandria, VA: Association for Supervision and

Curriculum Development

Lucas, A. M., P. J. McEwan, M. Ngware, M. Oketch (2014). Improving Early-Grade Literacy in East Africa: Experimental Evidence from Kenya and Uganda. *Journal of Policy Analysis and Management*. 33 (4)

Lundborg, Petter, Anton Nilsson, and Dan-Olof Rooth (2014). "Parental education and offspring 29 outcomes: evidence from the Swedish compulsory School Reform." *American Economic Journal: Applied Economics* 6, no. 1: 253-278.

Macdonald, K. A. D. and Vu, B. T. 2018. A randomized evaluation of a low-cost and highly scripted teaching method to improve basic early grade reading skills in Papua New Guinea (English). *World Bank Policy Research Working Paper Series* No. 8427. Washington, D.C. : World Bank Group

Magnuson, K. A., et al. (2007). Does prekindergarten improve school preparation and performance? *Economics of Education Review* 26(1): 33-51.

Naudeau, Sophie; Kataoka, Naoko; Valerio, Alexandria; Neuman, Michelle J. ; Elder, Leslie Kennedy. 2011. *Investing in young children: an early childhood development guide for policy dialogue and project preparation*. Washington, DC: World Bank.

Nakajima, N., Hasan, A., Jung H., Brinkman S., Pradhan M., Kinnell A. (2016) Investing in School Readiness: An analysis of the Cost Effectiveness of Early Childhood Education Pathways in Rural Indonesia. *World Bank Policy Research Working Paper Series* No. 7832.

National Institute for Child Health and Human Development 2000. *Report of the National Reading Panel. Teaching Children to Read: An Evidence-based Assessment of the Scientific Research Literature on Reading and its Implications for Reading Instruction*. (NIH Publication No. 00-4754). Washington, DC: National Institutes of Health

National Reading Panel. 2000. *Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction*. Washington, DC: National Institute of Child Health and Human Development.

Naudeau, S., N. Kataoka, A. Valerio, M. J. Neuman, L. Kennedy Elder (2010). Investing in Young People. An Early Childhood Development Guide for Policy Dialogue and Project Preparation. Conference Edition. Washington, D.C.: The World Bank.

Nores, M. and W. S. Barnett (2010). Benefits of early childhood interventions across the world: (Under) Investing in the very young. *Economics of Education Review*. 29:271-282.

O'Rourke, K., L. Howard-Grabman, and G. Seoane (1998). Impact of Community Organization of women in Perinatal Outcomes in Rural Boliva. *American Journal of Public Health*. 3 (1): 9-14

Piper, B. & Korda, M. (2011), *'Egra plus: Liberia. program evaluation report.'*, RTI International.

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.

Popova, A. Evans, D. Arancibia, V. 2016. Training Teachers on the Job : What Works and How to Measure It. *Policy Research Working Paper*, No. 7834 Washington, DC.: World Bank

Pradhan, M., D. Suryadarma, A. Beatty, M. Wong, A. Alisjahbana, A. Gaduh, R. P. Artha (2014). Improving educational quality through enhancing community participation: results from a randomized field experiment in Indonesia. *American Economic Journal: Applied Economics* Vol. 6, No. 2 (April 2014), pp. 105-126

Pressley, M., 1998. *Reading Instruction That Works: The Case for Balanced Teaching*. New York: The Guilford Press

Prost, A., T. Colbourn, N. Seward, K. Azad, A. Coomarasamy, A. Copas, T. A. J. Houweling, E. Fottrell, A. Kuddus, S. Lewycka, C. MacArthur, D. Manandhar, J. Morrison, C. Mwansambo, N. Nair, B. Nambiar, D. Osrin, C. Pagel, T. Phiri, A. Pulkki-BrännstrÖm, M. Rosato, J. Skordis-Worrall, N. Saville, N. S. More, B. Shrestha, P. Tripathy, A. Wilson, A. Costello (2013). Women's groups practising participatory learning and action to improve maternal and newborn health in low-resource settings: a systematic review and meta-analysis. *Lancet*, 381: 1736-46

Scarborough, H. S. 2002. Connecting Early Language and Literacy to Later Reading (Dis)abilities: Evidence, Theory, and Practice. In: Dickinson, D.K. and S.B. Neuman. *Handbook of Early Literacy Research (vol. 1)*. Edited by. New York: The Guilford Press: 97-110

Shonkoff, J. P. (2014). "Changing the Narrative for Early Childhood Investment." *JAMA Pediatrics* 168(2): 105-106.

Snow, C.E., Burns, M.S., and Griffin, P. 1998. *Preventing Reading Difficulties in Young Children*. National Academy Press, Washington, DC

Sprenger-Charolles, L. 2004. Linguistic Processes in Reading and Spelling: The Case of Alphabetic Writing Systems: English, French, German and Spanish. In: Nunes, T. and P. Bryant (Eds.) *Handbook of Children's Literacy*. Dordrecht, the Netherlands: Kluwer Academic Publishers: 43–66

Tauchmann, H (2012) "LEEBOUNDS: Stata module for estimating Lee (2009) treatment effect bounds," *Statistical Software Components* S457477, Boston College Department of Economics, revised 25 Jul 2013.

Toganivalu, D. 2008. Early Childhood Care and Education in the Pacific: Reflections of our past, our present and our future. In Puamau, P and Pene F. (Eds.) *Early Childhood Care and Education in the Pacific: The PRIDE Project*. Suva, Fiji: Institute of Education, University of the South Pacific.

Vegas, E. and L. Santibanez (2010*). The Promise of Early Childhood Development in Latin America and the Caribbean.* Washington, D.C.: The World Bank

Wolf, M. 2007. *Proust and the Squid: The Story and Science of the Reading Brain*. New York: Harper Collins

World Bank 2012a. *How well are Tongan children learning to read?* Washington, D.C.: The World Bank

World Bank 2012b. *How well are Ni-Vanuatu children learning to read in English?* Washington, D.C.: The World Bank

World Bank 2012c. *How well are Ni-Vanuatu children learning to read in French?* Washington, D.C.: The World Bank

World Bank 2013. *New Perspectives on Strengthening Government Capacity to Intervene for School Readiness in Samoa, Tonga and Vanuatu.* Washington, D.C.: The World Bank

World Bank 2014a. *East New Britain (ENB) Early Grade Reading Assessment (EGRA) Survey. 2012 Diagnostic Results Report.* Washington, D.C.: The World Bank

World Bank 2014b. *Madang Early Grade Reading Assessment (EGRA) Survey. 2011 Diagnostic Results Report.* Washington, D.C.: The World Bank

World Bank 2014c. *National Capital District (NCD) Early Grade Reading Assessment (EGRA) Survey. 2012 Diagnostic Results Report.* Washington, D.C.: The World Bank

World Bank 2014d. *Western Highlands Province Early Grade Reading Assessment (EGRA) Survey. 2013 Diagnostic Results Report.* Washington, D.C.: The World Bank

World Bank (2017a). *Tuvalu Early Grade Reading Assessment (TuEGRA): results report.*
Washington, D.C.: The World Bank

World Bank (2017b). *Kiribati Early Grade Reading Assessment (KiEGRA): results report.*
Washington, D.C.: The World Bank

Annex Figure 1. Density plots of TEGRA domain scores by sample and treatment status (in standard deviations)
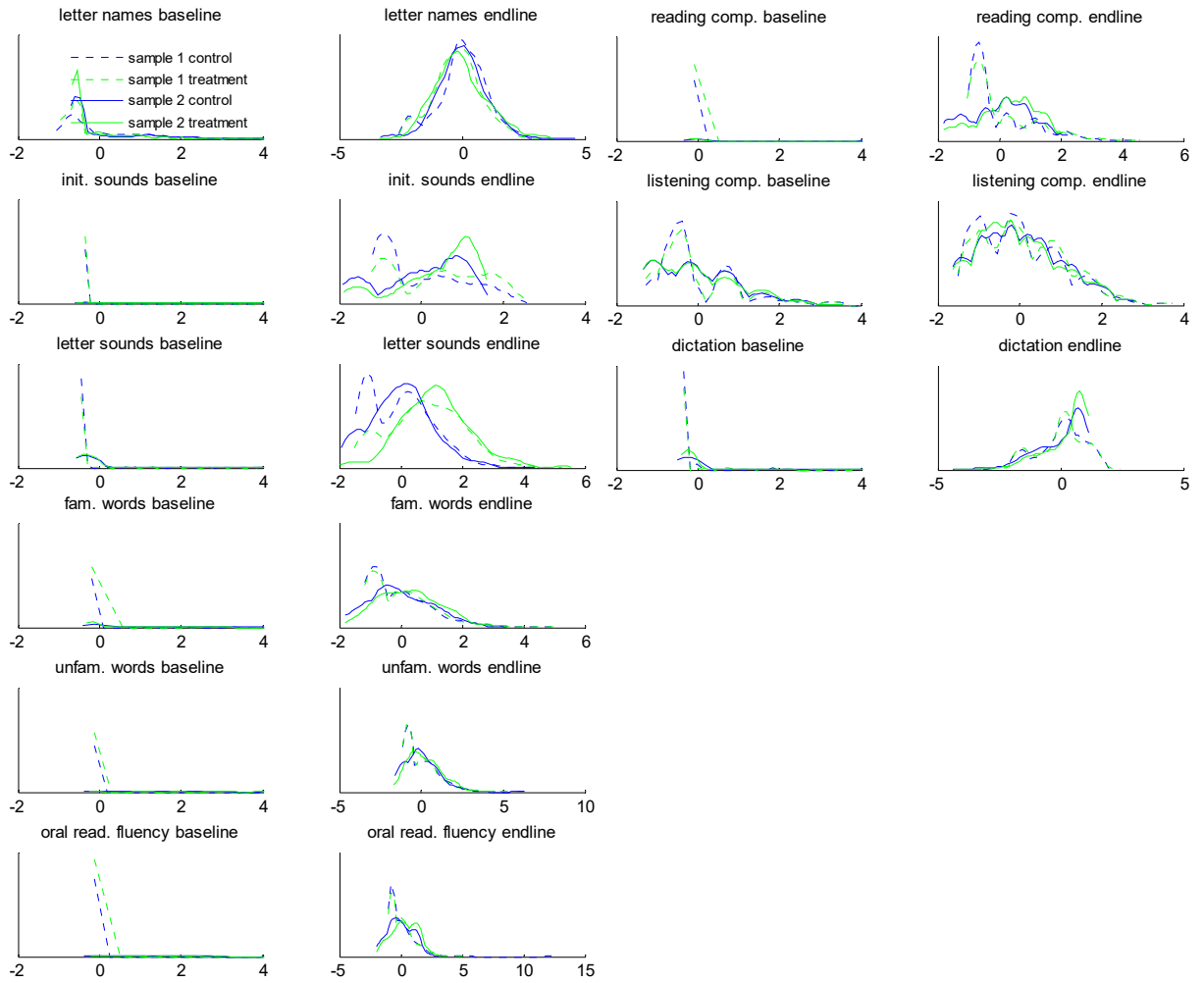
Annex Table 1. Estimates of equation (1)

	average TEHCI score (SD)	average literacy domains (SD)	home activities index	attending a CPBA	attending preschool
in a school readiness treatment community	0.04	0.07	0.03	0.21***	-0.02
	(0.06)	(0.05)	(0.03)	(0.02)	(0.02)
constant	-0.14***	-0.15***	0.69***	0.01***	0.34***
	(0.04)	(0.03)	(0.02)	(0)	(0.01)
observations	3429	3429	3426	3429	3429
R-square	0.00	0.00	0.00	0.10	0.00

Standard errors included in parentheses. For differences, statistical significance at the 1, 5 and 10 percent levels are denoted by ***, **, and *, respectively.

Annex Table 2. Estimates of equation (2)

	average TEHCI score (SD)	average literacy domains (SD)	home activities index	attending a CPBA	attending preschool
in a school readiness treatment (t) (effect on males)	0.08 (0.08)	0.17** (0.07)	0.05 (0.04)	0.2*** (0.03)	0.06** (0.03)
female (f)	0.12*** (0.04)	0.11*** (0.03)	0.04** (0.02)	-0.01** (0)	0.06*** (0.02)
mother completed high school (h)	0.64*** (0.05)	0.57*** (0.05)	0.2*** (0.03)	0 (0.01)	0.26*** (0.03)
treatment x female (tf)	0.12** (0.06)	0.03 (0.06)	-0.04* (0.02)	0.04** (0.02)	-0.04 (0.03)
treatment x mother comp. high school (th)	-0.15** (0.08)	-0.19*** (0.07)	-0.03 (0.04)	-0.01 (0.03)	-0.11*** (0.04)
female x mother comp. high school (fh)	-0.07 (0.05)	-0.03 (0.05)	-0.03 (0.02)	0.01* (0.01)	-0.03 (0.03)
treatment x female x mother comp. high school (tfh)	-0.07 (0.07)	-0.02 (0.07)	0.02 (0.02)	-0.02 (0.03)	0.02 (0.03)
constant	-0.6*** (0.05)	-0.58*** (0.04)	0.55*** (0.03)	0.01*** (0)	0.15*** (0.02)
observations	3429	3429	3426	3429	3429
R-square	0.07	0.06	0.06	0.10	0.04
effect on females with mothers who have not completed high school (t + tf)	0.2** (0.1)	0.19** (0.08)	0.01 (0.03)	0.24*** (0.03)	0.02 (0.04)
effect on males with mothers who have completed high school (t + th)	-0.07 (0.05)	-0.03 (0.05)	0.02 (0.03)	0.19*** (0.02)	-0.05** (0.03)
effect on females with mothers who have completed high school (t + tf + th + tfh)	-0.02 (0.07)	-0.02 (0.06)	0.01 (0.03)	0.21*** (0.02)	-0.07** (0.03)

Standard errors included in parentheses. For differences, statistical significance at the 1, 5 and 10 percent levels are denoted by ***, **, and *, respectively.

Annex Table 3. Effect of the school readiness intervention by domain and sub-population

	all	mothers without high school education		mothers with high school education	
		girls	boys	girls	boys
Verbal	0.02	0.02	-0.08	0.02	0
	(0.07)	(0.13)	(0.12)	(0.05)	(0.05)
Cultural & spiritual	0.17**	0.28***	0.24**	0.09	0.09
	(0.07)	(0.1)	(0.1)	(0.08)	(0.07)
Socio/emotional	0.01	0.17**	0.04	-0.03	-0.07
	(0.06)	(0.08)	(0.08)	(0.08)	(0.06)
Perseverance	-0.01	0.06	-0.03	0	-0.08
	(0.07)	(0.12)	(0.11)	(0.06)	(0.08)
Approaches to learning	0	0.28***	-0.08	-0.03	-0.1*
	(0.05)	(0.07)	(0.07)	(0.07)	(0.06)
Numeracy concepts	0.13***	0.26***	0.21***	0.05	0.04
	(0.05)	(0.08)	(0.06)	(0.05)	(0.05)
Reading	-0.03	0.04	0.05	-0.11	-0.09
	(0.06)	(0.09)	(0.07)	(0.07)	(0.06)
Writing	0.02	0.16**	0.17**	-0.05	-0.09**
	(0.04)	(0.07)	(0.07)	(0.05)	(0.04)
Physical	-0.05	0.01	-0.03	-0.06	-0.14*
	(0.07)	(0.1)	(0.11)	(0.08)	(0.08)

Standard errors included in parentheses. Statistical significance at the 1, 5 and 10 percent levels are denoted by ***, **, and *, respectively. Effects are estimated by equations (1) and (2).

Annex Table 4. Effect of the school readiness intervention by domain and sub-population on the TEHCI home activity measures

	all	mothers without high school education		mothers with high school education	
		girls	boys	girls	boys
reading	0.05*	0.09**	0.06	0.03	0
	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)
telling stories	0.02	0.03	0.02	-0.01	0.02
	(0.03)	(0.04)	(0.06)	(0.04)	(0.03)
singing songs	0.05*	0.02	0.11*	0.03	0.03
	(0.03)	(0.04)	(0.05)	(0.03)	(0.03)
taken outside the home	0.02	-0.02	0.01	0	0.04
	(0.04)	(0.05)	(0.06)	(0.04)	(0.04)
playing	0	-0.05	-0.03	-0.01	0.02
	(0.04)	(0.05)	(0.06)	(0.03)	(0.04)
naming or counting things	0.06*	0.03	0.11***	0.01	0.03
	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)

Standard errors included in parentheses. Statistical significance at the 1, 5 and 10 percent levels are denoted by ***, **, and *, respectively. Effects are estimated by equations (1) and (2).

Annex Table 5. Differences between attrition and non-attrition TEGRA sub-samples

	attrition sample	non-attrition sample	difference
1 year exposure sample			
average TEGRA score	-0.099	-0.026	-0.072***
	(0.017)	(0.014)	(0.021)
female	0.444	0.44	0.004
	(0.018)	(0.004)	(0.02)
background characteristics index	0.537	0.526	0.01
	(0.011)	(0.006)	(0.012)
2 years exposure sample			
average TEGRA score	-0.066	-0.034	-0.032
	(0.02)	(0.018)	(0.025)
female	0.375	0.423	-0.047*
	(0.017)	(0.009)	(0.024)
background characteristics index	0.51	0.537	-0.026**
	(0.009)	(0.007)	(0.011)

Standard errors included in parentheses. For differences, statistical significance at the 1, 5 and 10 percent levels are denoted by ***, **, and *, respectively.

Annex Table 6. Effect of CLRW on the proportion of teachers fully implementing teaching practices targeted by the intervention

	Treatment	Control	Difference
Year 2016			
phonemic awareness teaching practices	0.893	0.611	0.283***
	(0.008)	(0.021)	(0.021)
phonics teaching practices	0.92	0.52	0.399***
	(0.009)	(0.022)	(0.025)
reading teaching practices	0.956	0.621	0.334***
	(0.005)	(0.024)	(0.025)
writing teaching practices	0.952	0.391	0.561***
	(0.007)	(0.025)	(0.026)
sentence formation teaching practices	0.933	0.39	0.544***
	(0.01)	(0.023)	(0.025)
reading comprehension teaching practices	0.878	0.39	0.488***
	(0.011)	(0.029)	(0.031)
Year 2017			
phonemic awareness teaching practices	0.709	0.422	0.288***
	(0.03)	(0.03)	(0.05)
phonics teaching practices	0.908	0.641	0.267***
	(0.006)	(0.031)	(0.034)
reading teaching practices	0.909	0.725	0.184***
	(0.007)	(0.021)	(0.022)
writing teaching practices	0.88	0.451	0.429***
	(0.016)	(0.029)	(0.034)
sentence formation teaching practices	0.845	0.442	0.403***
	(0.017)	(0.024)	(0.029)
reading comprehension teaching practices	0.898	0.607	0.291***
	(0.01)	(0.027)	(0.029)

Standard errors included in parentheses. For differences, statistical significance at the 1, 5 and 10 percent levels are denoted by ***, **, and *, respectively.

Annex Table 7. Estimates of equations (4) and (5)

	equation (4)			equation (5)		
	sample 1	sample 2	sample 3	sample 1	sample 2	sample 3
baseline average TEGRA score	0.4***	0.33***	0.84***	0.39***	0.31***	0.81***
	(0.05)	(0.04)	(0.02)	(0.05)	(0.04)	(0.02)
CLRW treatment (effect on males for the gender model)	0.19***	0.33***	-0.09**	0.21***	0.38***	-0.04
	(0.04)	(0.06)	(0.04)	(0.04)	(0.08)	(0.05)
female				0.25***	0.32***	0.19***
				(0.03)	(0.06)	(0.03)
female x CLRW treatment				-0.03	-0.04	-0.07
				(0.04)	(0.08)	(0.05)
constant	-0.01	0.02	-0.02	-0.13***	-0.13**	-0.11***
	(0.03)	(0.04)	(0.03)	(0.02)	(0.05)	(0.03)
Observations	2199	818	862	2199	818	862
R-square	0.14	0.11	0.61	0.16	0.14	0.62
CLRW + female x CLRW (effect on females)				0.18***	0.34***	-0.11***
				(0.05)	(0.06)	(0.04)

Standard errors included in parentheses. Statistical significance at the 1, 5 and 10 percent levels are denoted by ***, **, and *, respectively.

Annex Table 8. Effect of CLRW on TEGRA reading domains (standard deviations)

	1 year exposure sample			2 year exposure sample			1 year after 2 years of exposure
	Effect, all	Lee bounds estimates		Effect, all	Lee bounds estimates		Effect, all
		Lower bound	Upper bound		Lower bound	Upper bound	
Letter names	-0.09**	-0.2***	-0.01	-0.05	-0.29***	0.21**	-0.01
	(0.04)	(0.05)	(0.07)	(0.08)	(0.07)	(0.08)	(0.06)
Initial sounds	0.43***	0.28***	0.49***	0.44***	0.35***	0.59***	0.1
	(0.05)	(0.04)	(0.04)	(0.08)	(0.07)	(0.17)	(0.07)
Letter sounds	0.85***	0.71***	0.94***	1.14***	0.89***	1.4***	0.37***
	(0.06)	(0.07)	(0.06)	(0.09)	(0.08)	(0.09)	(0.09)
Familiar words	0.07	-0.07	0.12**	0.3***	0.1	0.55***	-0.08**
	(0.05)	(0.05)	(0.05)	(0.08)	(0.07)	(0.08)	(0.03)
Unfam. words	0.01	-0.12**	0.06	0.23**	-0.04	0.44***	-0.02
	(0.05)	(0.05)	(0.05)	(0.09)	(0.07)	(0.08)	(0.04)
Oral reading fluency	0.07	-0.06	0.12**	0.36***	0.21***	0.64***	-0.03
	(0.05)	(0.05)	(0.05)	(0.07)	(0.06)	(0.09)	(0.04)
Reading comprehension	0.11**	-0.02	0.14***	0.31***	0.15***	0.46**	-0.12**
	(0.05)	(0.05)	(0.05)	(0.07)	(0.06)	(0.17)	(0.06)
Listening comprehension	0.1***	-0.01	0.17***	-0.03	-0.31***	0.15	-0.12*
	(0.04)	(0.05)	(0.05)	(0.07)	(0.07)	(0.1)	(0.07)
Dictation	0.03	-0.09**	0.09**	0.14**	0.04	0.42***	0.02
	(0.04)	(0.04)	(0.04)	(0.07)	(0.06)	(0.08)	(0.07)

Lee (2009) tight bounds are presented. The grouping variables for the 1 year exposure sample are cohort and gender, and for the 2 year exposure sample the grouping variable is gender. Gender is not used as a grouping variable for the bounds estimated for boys and girls. Standard errors included in parentheses. Statistical significance at the 1, 5 and 10 percent levels are denoted by ***, **, and *, respectively. Standard errors are estimated using Jackknife replication.

Annex Table 9. Estimates of equation (9)

	1 year of school readiness intervention sample		2 years of school readiness sample
	end of 1st grade average TEGRA score	end of 2nd grade average TEGRA score	end of 1st grade average TEGRA score
baseline average TEGRA score	0.49***	0.33***	0.34***
	(0.04)	(0.04)	(0.05)
% students in school from treatment comm. (sr) (effect on males in non-CLRW schools)	0.12	0.28*	0.01
	(0.13)	(0.15)	(0.1)
CLRW school (c)	0.27***	0.46***	0.19***
	(0.07)	(0.11)	(0.06)
female (f)	0.19***	0.38***	0.18***
	(0.06)	(0.08)	(0.06)
c x sr (difference in effect on males in CLRW and non-CLRW schools)	-0.04	-0.17	-0.1
	(0.18)	(0.21)	(0.12)
f x sr	0.05	0.02	0.28**
	(0.14)	(0.18)	(0.12)
f x c	-0.02	-0.11	0.08
	(0.1)	(0.11)	(0.07)
f x c x sr	0.05	-0.04	-0.28*
	(0.21)	(0.23)	(0.15)
cons	-0.14***	-0.27***	-0.14***
	(0.05)	(0.08)	(0.04)
N	1049	925	1140
r2	0.18	0.16	0.17
f x sr + sr (effect on females in non-CLRW schools)	0.17	0.3**	0.29**
	(0.15)	(0.12)	(0.13)
f x sr + sr + f x c x sr + c x sr (effect on females in CLRW schools)	0.18	0.09	-0.09
	(0.14)	(0.12)	(0.11)
f x c x sr + c x sr (difference in effect on females in CLRW and non-CLRW schools)	0.01	-0.21	-0.38**
	(0.21)	(0.17)	(0.17)
sr + c x sr (effect on males in CLRW schools)	0.08	0.1	-0.09
	(0.13)	(0.14)	(0.08)

Standard errors included in parentheses. Statistical significance at the 1, 5 and 10 percent levels are denoted by ***, **, and *, respectively.