School-based Management in Ugandan Primary Schools

Sarah Kabay, New York University

Elevate: Partners in Education

PRELIMINARY REPORT COMMENTS WELCOME

Abstract

Engaging local stakeholders in the management of primary schools is one of the most popular strategies for improving public education around the world. However, evaluations of these reforms have mixed results and underlying mechanisms are often unclear, raising questions as to why certain initiatives do or do not work and under what conditions. In this study, we contribute to the discourse on school-based management initiatives by advancing research on a particular intervention: a school scorecard program in which local actors are trained to monitor issues at their school. Originally evaluated in Uganda by Barr et al. (2012), a participatory version of the program was found to improve teacher attendance, student attendance, and test scores. A standard version of the program had smaller and statistically insignificant effects (Barr et al., 2012). The results of a behavioral game suggest that willingness to invest in the public good was a key mechanism behind the success of the participatory version (Barr et al, 2012). In a different region of Uganda, we implement the participatory version of the scorecard program, and combine process evaluation, qualitative research, and a quasi-experimental impact assessment to investigate the underlying mechanisms of impact and how the intervention interacts with the Ugandan educational context. Preliminary results find no impact on either teacher or student attendance, but a decrease in student transfer and school dropout. The process evaluation finds that schools focus on the role of parents and improving parent engagement.

Introduction

The results of a 2012 impact evaluation suggest that a school scorecard program might be one of the most cost-effective interventions for improving student learning in Uganda, and potentially in other country contexts as well (Barr et al., 2012; Zeitlin et al., 2011)[1]. The evaluation compared two variations of a primary school scorecard program: a "participatory" version, in which each school chose what issues to include on its scorecard, and a "standard" version, in which the research team chose the issues to be included. The participatory version decreased student and teacher absenteeism by 9 and 13 percent, respectively, and improved student test scores by 0.19 standard deviations. The standard version had smaller and statistically insignificant effects (Barr et al, 2012). The research team provided evidence to suggest that parents' sense of collective action was a key mechanism behind the success of the participatory version, but it is not completely clear how or why this treatment variation worked and the other did not (Barr et al, 2012). These findings align with much of the research on school-based management initiatives, which have mixed results and many outstanding questions as to how and why certain interventions work and under what conditions (Carr-Hill et al., 2016).

School scorecard programs and other school-based management initiatives represent the intersection of a number of prominent themes in education and development research: decentralization, accountability, community empowerment, and transparency. Though such interventions are very widespread and there is an extensive body of research focused on them, the underlying theory and even the definitions of these initiatives are often unclear. As an illustration, in one review of educational research, this scorecard program from Uganda was referred to as "participatory community-based monitoring" (J-PAL, 2017, p. 6), in another review it was included in a section on "incentive treatments" (McEwan, 2016, p. 372), and in a third review, it was described as a "low-stakes accountability program" (World Bank, 2018, p. 150). Furthermore, while the community aspect of this intervention is often emphasized, the scorecard committees were mainly comprised of school administrators and teachers. And though it is typically framed as an accountability intervention, the results of the original evaluation suggest that it might not be the power to hold someone accountable so much as the coordination and sense of solidarity between different stakeholders that determined the success of the participatory version.

In order to realize the potential of school-based management initiatives, clear and precise language and theorizing are needed. The conflation of different objectives and approaches to implementation cloud our understanding of what these different interventions are, how they work, and what they aim to achieve. The initial work of the Barr et al. (2012) research team to not just evaluate a scorecard program, but to compare different treatment variations and investigate underlying mechanisms of impact represents a unique opportunity to build on detailed theoretical and empirical analysis.

In the present study, I work with the non-governmental organization Elevate: Partners in Education to advance Barr et al.'s (2012) research on the school scorecard program. In a different region of Uganda, we implement the participatory version of the scorecard, along with a three-part research study including qualitative, process evaluation and impact assessment components. Our goals are first, to explore the potential for scale and sustainability of one of the most promising interventions for Ugandan primary schools, and second, to investigate the underlying mechanisms of the scorecard intervention, how

[1] Costs were estimated to be GBP 320 per school. Test scores were collected for just one grade. If costs are calculated just for this specific grade, including approximately 90 students, a 0.1 sd improvement in test scores costs GBP 3.60. As a whole school intervention, it is highly likely that other grades were also impacted, in which case the costs would decrease significantly.

it interacts with the Ugandan context, and how building on the evidence based for this particular intervention might contribute to understanding of school-based management more broadly.

Several aspects of the research project are on-going. Both the qualitative study and the process evaluation will continue through the end of the 2019 school year, in November. Additional analysis is planned for the impact assessment, and we are in the process of obtaining data from Uganda's National Assessment of Education Progress (NAPE) exams, which the National Examinations Board plans to release in June 2019.

In this paper, I accordingly present only preliminary results. I begin by providing some background information on school-based management initiatives, the Ugandan context, and the original study conducted by Barr et al. (2012). I then describe the methods and sample of the current project, and preliminary findings from the process evaluation and impact assessment. I end with a few preliminary conclusions.

Background

School-Based Management

School-based management (SBM) can be broadly defined as the devolution of power and responsibility from a central authority to individual schools, and as a focus on the individual school as the primary unit of improvement (Malen, Ogawa, & Kranz, 1990). It has been one of the most popular education reforms of the past half century. First implemented in a few high-income countries in the 1970s and 1980s, SBM became "a centerpiece in the broader school restructuring agenda of the 1990s" (Leithwood & Menzies, 1998, p. 326). More recently, SBM has expanded into low-income countries, partially due to the support of international development agencies such as the U.S. Agency for International Development, the UK Department for International Development, and the World Bank, which, for example, devoted $1.7 billion dollars to SBM reforms between 2000 and 2006, representing 18% of the Bank's total education financing (Barrera-Osorio et al., 2009; Carr-Hill et al., 2016). At this point it would be hard to find an education system untouched by some form of SBM reform.

A significant body of research has accompanied this interest and investment in SBM, with mixed results. A review of SBM studies from 1998 analyzed over eighty different empirical studies; 72 different effects were reported, of which 45 were positive and 27 were neutral or negative (and mostly negative) (Leithwood & Menzies, 1998). A 2016 review of SBM in low- and middle-income countries included 26 impact studies and 9 non-causal studies. The review identified six outcomes for which there were enough data to conduct meta-analysis, such as dropout, student learning, and teacher attendance. The authors identified 19 estimates for student learning, of which one was negative and statistically significant, five were positive and significant, and the remainder were neutral or insignificant (Carr-Hill et al., 2016).

Though the rigor of research on SBM has vastly improved in the 18 years between these two research reviews, the mixed results remain difficult to interpret. First of all, because SBM refers to a wide variety of different reforms and interventions, one challenge is to simply categorize them. Two important dimensions are *what* and *who*, as in *what* power or responsibility is being transferred and to *whom* is it being transferred (Barrera-Osorio et at., 2009). One suggested categorization focuses on *who*, defining four broad forms of SBM in relation to the primary agent to whom power is transferred: administrative (principal), professional (teachers), community, and balanced control (Leithwood & Menzies, 1998; Barrera-Osorio, 2009). An alternative approach is to identify specific models of SBM. One study estimates that in the United States alone there are more than 800 different models of SBM (Rowan, Bamburn & Barnes, 2004).

Synthesizing the results of SBM initiatives can accordingly be comparing apples to oranges, but recent reviews of the research suggest a few cross-cutting themes. In some cases, these themes echo the findings of earlier research. The aforementioned review from 1998 concluded that "significant improvements in student learning fundamentally depend on the widespread implementation of more effective teaching practices," which is primarily "a problem of teacher learning, not a problem of organization or structure," and, accordingly, not a problem that SBM can necessarily solve (Leithwood & Menzies, 1998, p. 342). Similarly, the 2018 World Development Report summarized its section on school management by stating that school management "will improve learning only if it directly improves the quality of teacher learner interactions," which would be unusual for many SBM interventions, such as community monitoring, because "parents are rarely in the classroom, and even when they are, they cannot necessarily identify good classroom practice" (World Bank, 2018, p. 149).

These two research reviews draw attention to the fact that, in relation to student learning, SBM is a relatively distal intervention. Multiple mechanisms need to be in place in order for SBM to affect test scores or other academic outcomes. A review of community accountability and empowerment interventions, for example, identified four intermediate levels of outcomes between the specific activities of SBM and the ultimate outcome of improved student learning (Westhrope et al., 2014).

And yet, a number of SBM interventions *have* positively impacted student learning. One theory is that SBM will only work in relatively advantaged communities, where parents or other community members have a certain level of human capital, as was found in a study from The Gambia (Blimpo, Evans & Lahire, 2015). Another theory is that SBM interventions depend upon collective action, and factors that support efforts to coordinate might represent an enabling condition for success of SBM interventions. This argument is supported by Barr et al. in relation to ethnic heterogeneity. Along with other research from Uganda (Habyarimana et al, 2007) Barr et al. hypothesize that ethnically homogenous communities are better able to address collective action challenges than more diverse ones, and are therefore better able to implement the scorecard intervention.

A number of issues remain unclear: what types of SMB reforms are effective and why? And how do general contextual factors translate into specific actions and impacts? More specifically, what might parents with greater human capital do in the context of scorecard initiatives that other parents might not? And what norms, beliefs, or activities enable ethnically homogenous communities to implement SBM reforms that might be absent in more diverse communities? Some theoretical work exists: Westhorp et al.'s (2014) review of community accountability and empowerment interventions enumerates 11 possible mechanisms through which such interventions can improve educational outcomes, along with 28 different elements of context that seem to affect the efficacy of these interventions, but the authors recognize that many of these pathways are vague and hypothetical. The report concludes that future research needs to more explicitly investigate theories of change (Westhorp et al., 2014). Similarly, the 2016 review of SBM interventions argues that limited evidence prevents any very robust conclusions on potential barriers or enablers of effects, or on the conditions necessary for positive impact (Carr-Hill, 2016).

Original Study

The Barr et al. (2012) evaluation represents an important step towards understanding mechanisms and contexts for change, as its research design includes two features that enable deeper inference about the scorecard and its impact. First, the authors compared two different treatment variations (participatory and standard) in addition to a pure control which received no intervention. Second, they conducted a behavioral game designed to measure willingness to cooperate and invest in the public good. Comparing differences in the implementation and impact of the two treatment variations, and aligning these findings

with the results of the behavioral game, helps to shed light on the underlying mechanisms of the scorecard and enabling conditions.

To briefly describe the intervention, in all treatment schools, a general meeting is held with a school's administration, teachers, parents, and the School Management Committee (SMC). During this meeting, the scorecard program is described and participants are asked to elect nine individuals: three SMC members, three parents, and three teachers. These nine individuals join the headteacher and two pupil representatives to form the 12 members of the scorecard committee. The scorecard program can accordingly be classified as "balanced control" SBM, in which administrators, professionals, and community members are all involved.[2]

Following the initial general meeting, the scorecard committee members are given an additional two days of training on how to implement the scorecard. Ultimately, the committee members each individually visit the school and monitor the issues included on the scorecard. At the end of each term, the committee members meet to reconcile their results into a single "consensus" scorecard, and share this information with the District Education Office and with the school community at the next general Parent Teacher Association meeting.

The two treatment variations each follow this same process, the only difference between them being how items are selected to be included on the scorecard. In the "standard" variation of the scorecard, the study's researchers work with implementing NGOs and officials from the Ministry of Education to select issues to be included on the scorecard. In the "participatory" variation of the scorecard, at the same general meeting where participants elect the members of the scorecard committee, they also elect the issues to be included on the scorecard. The participatory scorecard accordingly varies by school, while the standard scorecard is the same for all schools in that treatment arm.

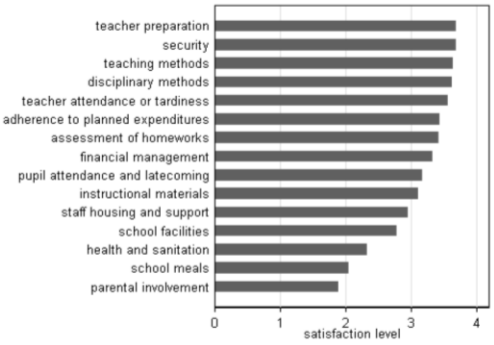
The Barr et al. (2012) study was conducted in 100 rural government primary schools spread across four different districts. Each treatment arm included 30 schools and the remaining 40 schools served as the control group. The study included four primary sources of data: NAPE exams and random spot-checks to measure teacher and student attendance, which were collection from all schools, and data from the scorecards and behavioral game, which were collected only in treatment schools. Schools which implemented the participatory version of the scorecard improved teacher and student attendance and student test scores. Schools with the standard version of the scorecard had smaller impacts, which were not statistically significant (Barr et al., 2012).

A first line of inquiry is to determine if the content on the scorecard differed between the two treatments. It is possible that school communities know better than the researchers which issues are most important to their school. For schools with the standard scorecard, Barr et al. (2012) reported the average satisfaction level across schools for each issue included on the scorecard (from 1 = "Very unsatisfactory" to 5 = "Very good") (Figure 1 A), as well as the issues that were most frequently included in participatory scorecards (Figure 1 B).

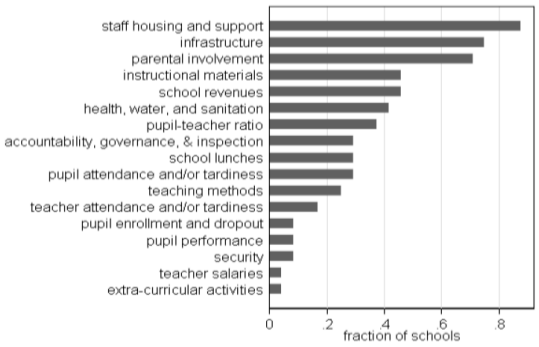
[2] An outstanding question is who is considered a community member, given how often this intervention is simply described as community-based (JPAL, 2017; Carr-Hill, 2014), when in actuality, of the twelve members of the committee, four are salaried professionals, and an additional three are officially appointed volunteers or local government officials.

Figure 1.

A. Standard Scorecard



B. Participatory Scorecard



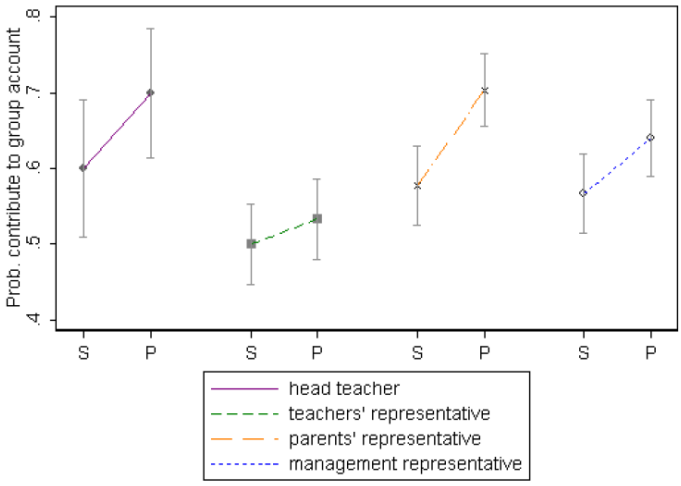
From Barr et al., 2012

There is observable overlap and congruity between the standard and participatory scorecards. Issues that were most frequently featured in participatory scorecards (parental involvement, staff housing and support) were among the issues that were rated as most unsatisfactory on the standard scorecard. There are some issues that are unique to either the standard or participatory scorecards, but generally, both scorecards focused on a wide range of different issues.

Though high rates of teacher absenteeism were one of the key motivations behind this study, less than 10% of the 50 schools with the participatory scorecard included it on their scorecards. Schools in the standard treatment rated it as one of the issues with which they were most satisfied. Given that teacher attendance improved in schools in the participatory arm, it is notable that these improvements occurred even when schools were not monitoring these issues. Conversely, teacher attendance did not improve in schools in the standard arm, which were both monitoring and satisfied with teacher attendance. In addition, a few of the issues that were most popular in the participatory treatment, namely staff housing and infrastructure, did not change over the course of the study, in either treatment. Though described as an "information-for-accountability" intervention, it does not seem as though the participatory scorecard program worked by monitoring and communicating information on specific issues (Barr et al., 2012).

The results of the behavioral game provide an alternative hypothesis. In both treatment groups, at the conclusion of the training, researchers had the members of the scorecard committee participate in a binary public goods game. The Voluntary Contributions Mechanism, following implementation as by Cardenas and Jaramillo (2007), is designed to measure individual's willingness to invest in the public good. Across all four stakeholders, Barr et al. find that the impact of the participatory treatment was an 8 percentage point increase in the probability of contribution to the public good. Differentiating by stakeholder, Barr et al. (2012) find a substantively and significantly stronger impact on parents. Parents in the participatory treatment were 16 percentage points more likely to contribute to the public good than parents in the standard treatment (Figure 2.)

Figure 2. Results of VCM Behavioral Game: Standard Treatment (S) vs. Participatory Treatment (P)



From Barr et al., 2012

Barr et al. (2012) further advance this theory by including information on the ethnic background of parents. They find that a school's ethnic homogeneity is positively associated with contributions to the public good, and also that an interaction term between the participatory treatment and their measure of ethnic homogeneity (the percent of parents in the school that identify as the dominant ethnicity) was positive and statistically significant (Barr et al., 2012).

Barr et al. (2012) conclude that individuals' willingness to act in the interest of the community, as measured by the VCM game, served as a mediating mechanism for the participatory scorecard. Across three key outcomes – teacher attendance, student attendance, and student test scores – the participatory scorecard outperformed both the control schools and schools with the standardized scorecard. Compared to the standard version, the participatory scorecard "better engaged the community in a process of discussing school goals, constraints, and progress" (Barr et al., 2012, p. 30).

While this study advances our understanding of school management by highlighting the importance of collective action and the role of ethnic homogeneity as an enabler of collective action, it also raises a number of questions. First, if there was no correlation between the issues on the scorecard and the issues for which outcomes were observed, what role did the monitoring play? Over the two years of the study, the vast majority of program activity would consist of monitoring the issues on the scorecard and then communicating that information to other stakeholders. How did that activity translate into impacts if the issues the scorecard committees were monitoring were not the issues for which there were impacts?

Second, what role did parents play? Comprising only a quarter of the scorecard committee, but experiencing the most dramatic change in collective action, Barr et al. write that parents are "the principal focus in local-accountability interventions" (p. 27). How does this relate to the finding that in the standard scorecard, parent involvement was the issue rated as most unsatisfactory? And finally, teachers had the lowest levels of collective action of all the stakeholders, and yet over 90% of participatory schools were monitoring staff housing and support. In the impact report from this study, the authors note that there were no observed changes in staff housing and support (Zeitlin et al., 2011). Why then did teacher absenteeism decrease?

Ugandan Context

It is important to contextualize some of these findings within the education system and broader education landscape of Uganda. Like many low-income countries, Uganda has struggled with both access to and the quality of its primary education. The results of numerous studies and citizen-led assessment initiatives indicate that the majority of children are far behind in basic literacy and numeracy (Uwezo, 2015). Teacher and student absenteeism are widespread, and there are high rates of grade repetition and dropout (Education Policy and Data Center, 2011). The country has a growing private primary education sector and high rates of transfer between schools, both public and private (Kabay, 2019).

Decentralization has long been a key feature of education policy in Uganda. When the current government came to power in the late 1980s, decentralizing public services from the national level to the district level was one of its primary reforms (Namukas & Buye, 2009). School-based management in Uganda predates even this initiative, as school management committees (SMC) were first established in 1969 (Najjumba, Habyarimana & Bunjo, 2013). Originally comprised of local government officials, members of the school's founding body, and a teacher, parent, boy and girl, SMCs have four primary responsibilities: financial management, infrastructure management, management of students and staff, and conflict resolution (Najjumba, Habyarimana & Bunjo, 2013).

Research on SMCs in Uganda has found that their members receive little training and are often unaware of their roles. In general, school management practices and accountability are weak (Najjumba, Habyarimana & Bunjo, 2013). A key challenge concerns the role of parents. One theory argues that the government's 1997 policy of abolishing tuition payments undermined local accountability (Dauda, 2004). In qualitative research, parents have explained that they no longer have a say in primary education, as it is a service for which they are not paying (Marphatia et al., 2010). In fact, Uganda's Universal Primary Education is a cost sharing policy in which parents do have some financial responsibility, but miscommunication of this policy has generated significant confusion and distanced parents from schools (Avenstrup, 2004). From the perspective of headteachers, 97 percent of school-level problems relate to parent and community factors, as opposed to in-school factors such as teachers or curriculum (Najjumba, Habyarimana & Bunjo, 2013). Similarly, a survey of teachers from 2011 found that, aside from teachers' salaries, teachers felt that lack of parent and community support for primary education was the most important challenge facing Ugandan primary schools (Kabay, 2016). Barr et al.'s findings that scorecard committees were dissatisfied with parents' involvement is consistent with this literature.

It is possible, though, that parents are an easy scapegoat for educational problems in Uganda. In fact, individual households contribute more to educational expenditure than the government. Data released by UNESCO's Institute for Statistics finds that Ugandan households cover 57% of total education expenditure (UNESCO, 2018). Better understanding the division of roles and responsibilities among different stakeholders, and how those roles and responsibilities are perceived, could significantly contribute to understanding school-based management in Uganda, and accordingly the school scorecard program. This issue is one of the key themes of interest in the current study.

Teachers are also an important stakeholder. During the study period of the Barr et al. evaulation, there were at least two nationwide teacher strikes (Daily Monitor, 2011; Kiyaga & Ssenkabirwa, 2011). This might explain why teachers had the lowest levels of contribution in the VCM game in Barr's research. It might also be related to why participatory schools were monitoring staff housing. Though the government's response to the strikes was to threaten and even imprison teachers and union leaders, the striking teachers did receive some support. For example, an editorial in a national newspaper at the time had the headline "Government should listen to teachers" (Daily Monitor, 2011). Given that teachers were earning as little as 70 USD a month, school communities might have chosen teacher accommodation on

their scorecard as a way in which to help compensate their teachers. Rather than holding teachers accountable, it seems school communities might have been working to support or otherwise demonstrate their solidarity with teachers. The role of teachers is a second focal point of the current study.

Present Study

Implementing Organization

Elevate: Partners in Education is a nonprofit organization focused on improving primary education in Uganda. Elevate is a small NGO, with only five full time employees. In the past, Elevate has implemented traditional charitable activities, such as paying for children's school fees and supporting local NGOs. In an effort to become more evidence-based and rigorous in its work to improve primary education, Elevate reviewed different educational interventions and determined that the participatory school scorecard would be the most cost-effective way in which to improve student learning and capitalize on Elevate's strengths as an organization. Elevate began by piloting the participatory version of the scorecard program in 2014 in 13 schools across Uganda using the same program implementations as in Barr et al. (2012).[3]

Elevate's pilot caught the attention of the District Education Officer (DEO) of Mukono, a district in Uganda's central region, who was interested in bringing the scorecard program to his district. This presented the opportunity for Elevate to work in more schools and also conduct research. As a small organization with a limited budget, Elevate was not in a position to immediately implement at scale, but by building off of the original study by Barr et al. and learning more about the intervention, Elevate could strategically operate in the space between the initial RCT and large scale implementation. With such experience investigating the scorecard program's potential for sustainability and scale, Elevate aims to later serve in an advisory or training role for at-scale implementation.

In Mukono, Elevate wanted to foreground the question of scale in two ways, first by implementing at the subcounty level (the governing unit below district) and second by experimenting with the profile of a trainer. In relation to the first, rather than work just at the level of the individual school, Elevate wanted to enable schools to collaborate and work together. If the scorecard program were to be implemented in hundreds of schools, such coordination could assist with at-scale implementation and Elevate wanted to explore this possibility by fostering connections between schools located within the same subcounty. In relation to the second, implementation in the Barr et al. evaluation used Central Coordinating Tutors (CCTs) to conduct the three-day trainings and facilitate meetings with scorecard committees. In Uganda, CCTs serve a general support role for primary education, organized through the teacher training colleges. Concern for the over-burdening of this workforce motivated Elevate to test two other possible trainers: Associate Assessors and social workers. Ultimately, Elevate would work with one CCT, two Associate Assessors, and two social workers to implement the scorecard program in Mukono.

Elevate accordingly had three interrelated goals for implementing the participatory school scorecard program: explore the intervention's potential for sustainability and scale, investigate possible

[3] Throughout Elevate's work, the Barr et al. research team has been extraordinarily helpful in answering questions and sharing both research and implementation materials. Elevate was able to use the same program manual to guide implementation. The study's data collection instruments similarly informed Elevate's research tools. We were able to speak or meet with four different members of the Barr et al. evaluation, including the point person on the ground. This support was invaluable.

mechanisms of impact, and develop Elevate's expertise as an implementer of the scorecard program. The associated research was designed to align with these objectives.

Research Design

The present study includes three components: qualitative research, process evaluation, and impact assessment. Each component is defined by its own data and methods, but the three are intentionally complementary and interrelated. Together, they address Elevate's goal to understand implementation of the scorecard intervention, and also contribute to the broader discourse on SBM. Individually, each component focuses on a few specific topics and research questions. One important distinction is that the qualitative research and process evaluation data were collected by the Elevate team, while the impact assessment data were collected by independent enumerators.

Qualitative Study

The qualitative study is broadly focused on understanding the Ugandan context and how the scorecard intervention interacted with existing dynamics and issues within this context. Following Rubin and Rubin's (2005) method for responsive interviewing, the qualitative study consists of multiple waves of semi-structured interviews. We use the qualitative research to uncover new issues in a Grounded Theory approach (Charmaz, 2001), and also to investigate three topics we believed to be important to understanding how the scorecard intervention operates.

The three focal topics of the qualitative research are parents, teachers, and collective action. As previously described, many Ugandan actors believe the behavior of parents is the most serious challenge facing primary education. The first objective of the qualitative research is to interrogate this narrative. We aim to more deeply understand the expectations and perspectives of school administrators and parents in relation to the role of parents in primary education, both in general and in relation to implementation of the scorecard intervention. This objective was in part inspired by the results of the Barr et al. (2012) study which suggest that parents' sense of collective action was a key driver of impact. The second objective is to similarly investigate the role of teachers. This topic also relates to the Barr et al. (2012) study, and the fact that 90 percent of participatory schools were monitoring teachers' accommodation, and was also motivated by some early findings of the present study. And finally, in our qualitative research we investigate the theory that the scorecard operates through an improved sense of collective action. This includes discussing research participants' understanding of the VCM behavioral game, encouraging them to describe and explain the scorecard intervention and their experience of it, and to ask about the relationships between different stakeholders in education.

Key research questions for the qualitative study include how do different stakeholders perceive the scorecard intervention? What issues define the role of teachers in Ugandan primary education? And why are parents believed to be a key challenge in Ugandan primary education? The qualitative study includes two methodological approaches: facilitator interviews and researcher interviews. The facilitator interviews were conducted by the CCT, Associate Assessors, and Social Workers Elevate contracted to directly implement the scorecard program in schools. Each facilitator used a standard, open-ended survey to conduct a key informant interview at each of their schools. They filled in a blank form designed for this purpose, hand writing the response to questions such as "What role does the scorecard program play in improving education standards?" These interviews have been completed.

In the researcher interviews, Elevate's Program Manager, Elevate's Monitoring and Evaluation Officer, and I conducted more extensive, semi-structured interviews. Schools were randomly selected to be included from both the control and treatment groups, and individual interviews were conducted with

teachers, headteachers, parents, and school management committee members at each school. Interviews ranged in length from half an hour to an hour. Interviews were recorded and transcribed. At this point, we have conducted four rounds of qualitative data collection, totaling 37 interviews. During the months of June and July 2019, we plan to conduct three more rounds of interviews. In line with the responsive interviewing approach, we briefly analyze and discuss interview transcripts after each round of interviews and use this work to inform subsequent rounds of interviews. We have not yet conducted formal coding and analysis.

Process Evaluation

A central part of the study is to take an intervention science approach to Elevate's implementation of the scorecard, which we call the process evaluation (Forman et al., 2013). Generally, the process evaluation aims to understand barriers to implementation, identify core intervention components, and theorize mechanisms of impact and underlying theories of change. Weekly discussions with the Elevate team were used to explore these issues, and hypothesize enabling and constraining factors affecting implementation. Elevate collected all data for the process evaluation, aiming to identify and document all program activity, including the initial and on-going activity of the Elevate team, as well as intermediate steps and actions both attempted and completed by schools.

The process evaluation includes two primary sources of data. The first is the "program activity tracker." For each school, the issues elected to be on the scorecard, the indicators schools used to monitor these issues, the person responsible for completing each activity, and feedback on whether it was successfully completed were all recorded. These data were all collected by Elevate facilitators, the CCT, Associate Assessors, and Social Workers who were each in charge of approximately ten schools. Data were then collated and inputted by Elevate's Monitoring and Evaluation Officer.

The second data source is the "school rating" exercise. First, as a team Elevate brainstormed criteria which could be used to evaluate schools' implementation of the scorecard. At two points over the course of the two-year implementation period, the Elevate team rated the status of program implementation at each school as either strong, average, or weak. This exercise resulted in two sets of findings, first the theory and criteria that Elevate believed defined program implementation, and second rates assigned to each school.

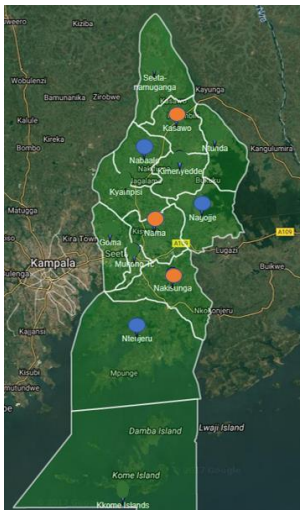
Impact Assessment

While Elevate intended to conduct a full randomized controlled trial, only certain CCTs and Associate Assessors were available to participate as trainers in the scorecard program, which stipulated that the intervention be implemented in two specific subcounties. Elevate accordingly decided to use a matched / quasi-experimental design. Elevate consulted with multiple education experts and officials familiar with Mukono District in order to create three pairs of matched subcounties. Factors considered in this matching process included urbanicity, economic activity, education performance, and major infrastructure such as roads. In two pairs, assignment to treatment was determined by the availability of the CCT or Associate Assessors. In (only) the third pair, with the social workers, it was possible to randomize which subcounty within the pair was assigned to treatment.

Within each subcounty, Elevate randomly selected schools to participate in the research until the study included 100 schools, equally distributed across treatment and control, and across subcounty, to the extent possible. On average, each subcounty included 17 schools (range from 13 to 20). Fifty schools were in treatment subcounties and fifty in control. Using the PowerUp tool for multilevel model statistical power analysis (Dong & Maynard, 2013), and specifications for R squared and ICC values from the

baseline Uwezo English scores, we have 80% power to detect a 0.344 sd minimum detectable effect on student-level outcomes. Although far from ideal, we believe the research design still enables us to make certain inferences, taking into account its limitations of statistical power and internal validity.

Figure 3. Subcounties in Mukono: Treatment in Red, Control in Blue



The baseline was conducted at the end of 2016 and beginning of 2017, including measures for approximately two thousand children and one thousand teachers. Program implementation in the 50 treatment schools began immediately after and continued throughout the 2017 and 2018 school years. The endline was conducted at the end of 2018. Quantitative data collection was completed at the end of March 2019, after a final spot-check for teacher and pupil attendance. Qualitative data collection will continue through July 2019 and the process evaluation will continue through the end of 2019.

Measures. The baseline included three data collection instruments: a school survey conducted with the headteacher, an individual assessment of children's learning, and a random spot-check to assess teacher and student attendance. The school survey focused on general school characteristics, such as infrastructure and student enrollment, along with a number of questions about the headteacher's perspective, asking the headteacher what he/she thinks is the biggest problem facing the school or whether he/she feels supported by parents, etc.

For the baseline individual assessment of children's learning, a sample of children were randomly selected from the second-grade classroom in 2016 and the third-grade classroom in 2017. The Uwezo assessment was administered for each sampled student. Inspired by the ASER assessment initiative conducted by Pratham in India, the Twaweza initiative was started in East Africa, conducting citizen-led assessments in Kenya, Uganda, and Twaweza. Uwezo represents the Ugandan arm of Twaweza and developed basic assessments of early literacy and numeracy according to second-grade standards. We used the Uwezo test booklet from the 2015 year. Because schools varied greatly in size, the number of children sampled per school varied widely. At most we sampled 30 children per school, in the smallest school, there was only one child in second-grade. Overall, the baseline sample included 2,117 children.

The Uwezo assessment represents a significant departure from the test scores used in the original study. Uwezo directly assesses basic literacy and numeracy skills, such as identifying letters, reading individual words, and performing basic addition and subtraction. Barr et al. used the Ugandan national standardized exams, the NAPE. We are fortunate in that the two most recent waves of the NAPE perfectly correspond with our study timeline (2016 and 2018), though we are not sure how much overlap there is

between the specific schools in our study and the schools in which the NAPE was conducted. We are in the process of obtaining the NAPE data from Uganda's National Examinations Board. The NAPE data might also enable additional analyses, such as comparing the scores of schools within the same subcounty that did and did not receive the scorecard (essentially a randomized comparison), or the scores of additional districts in Mukono.

For the attendance spot-check, an enumerator visited the school unannounced and collected data on teacher and student attendance. In the third-grade classrooms, the enumerator physically counted the number of children present in the classroom, as well as the number of children wearing a uniform and shoes. For total school enrollment and attendance, the enumerator requested the information from the headteacher. For teacher attendance, enumerators obtained a staff list from the headteacher, and then would walk throughout the school compound directly observing the location (or absence) of each individual teacher.

For the endline, we conducted a follow- up version of the school survey with the headteacher, asking similar questions as during the baseline. For the Uwezo assessment, we worked to track the exact children who were assessed during the baseline. We were able to find and assess 1,657 children, or 78% of the baseline sample. This follow-up sample was equally balanced across treatment and control. For this tracked sample, children's age, grade, and gender were collected along with the Uwezo assessment, and a few questions such as when was the last time a parent had visited their school.

During the endline, we also conducted the same Voluntary Contributions Mechanism game as was conducted in the original study by Barr et al. (2012). Although we used the same script and protocol, our implementation of the game differed from the original study in a few key ways. First, in the original study, the game was conducted immediately after the scorecard training session. The authors accordingly categorize the results of the game as a mediator. We conducted the game during the endline -- it might therefore be more appropriately categorized as an outcome. Second, in the original study, the game was conducted only in the treatment arms, as it was played with members of the scorecard in both the standard and participatory treatments. We did not implement the standard version of the scorecard, and so had to compare schools with and without a scorecard committee. Without an equivalent comparison group, we had to improvise. In the control group, we requested the participation of key leaders amongst the SMC, parents, and teachers to approximate the members of the scorecard committee. Given that we conducted the game during the endline data collection, all other setting and conditions were the same across treatment and control.

In addition to the behavioral game, we also asked all respondents a series of questions tailored to their role (parent, SMC, teacher, headteacher). These questions were inspired by early findings from the qualitative study and process evaluation. They were intended to represent various theories for how the scorecard program worked and what might have supported/hindered program implementation. For example, in asking teachers if they feel supported by their community, we are interested in determining if the scorecard intervention might have worked through improving the relationship between teachers and community members, and if this might be related to teachers' sense of collective action. All of these questions were framed in relation to a Likert scale (Likert, 1932) from Strongly Disagree to Strong Agree. And as in Barr et al. (2012), we collected information on respondents' ethnicity.

The primary research questions for the evaluation were did Elevate's implementation of the participatory scorecard program generate the same pattern of results as in Barr et al.'s (2012) evaluation? And were any other educational outcomes effected?

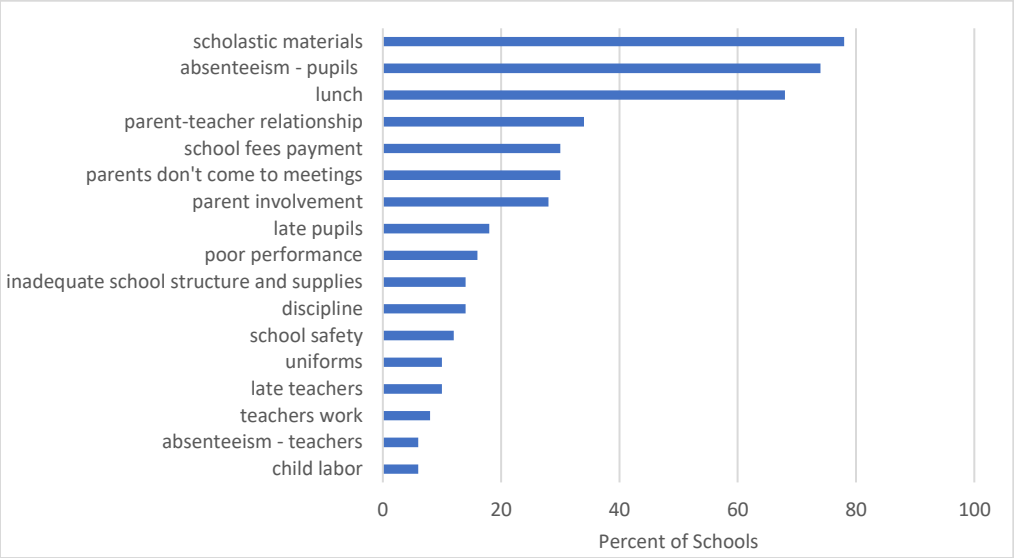
Preliminary Results

As of June 2019, all data for the impact assessment have been collected. We are waiting only to gain access to the NAPE data, which the Ugandan Examinations Board has reported will release in June. Initial analysis of the individual student assessment (Uwezo scores), school survey, and teacher and pupil attendance has been completed. I am able to report preliminary results for these data sources, but analysis is on-going. Data collection for both the qualitative study and process evaluation is also on-going. Two more rounds of qualitative data will be collected in June and July of 2019 and the process evaluation will continue throughout 2019, as Elevate tests the sustainability of the intervention and schools' ability to run the scorecard with minimal support from Elevate. In this paper I report preliminary findings from the process evaluation and impact assessment.

Process Evaluation

Starting with the program activity tracking, participants in the initial general meeting of the scorecard intervention voted on five issues to include on their school's scorecard. Figure 4. reports the percentage of schools that included specific issues on their scorecard.

Figure 4. Issues Included on the Scorecard



Of the seven most popular issues, six directly refer to parents, and the seventh, pupil absenteeism, could easily be seen to indirectly relate to parents. In Uganda's cost sharing policy, parents are responsible for providing children with scholastic materials and lunch. School fees are negotiated by each school individually, but again, are the responsibility of the parent. While some scorecards simply described an issue as, for instance, "Pupils have no lunch at school," others would most explicitly identify the issue as "Parents do not pay lunch for their children." Consistent with other research from Uganda, the scorecards highlight the perception that parents are a critical challenge in Uganda's primary education.

The Elevate implementation team documented for each issue at each school what indicators scorecard committees used to monitor the issue, what steps the school planned to take in order to address the issue, who was responsible for each step, and feedback on how implementing the step went. As an illustration, one of the schools that included the issue of school fees payment on their scorecard decided that the headteacher would write letters to remind parents who had not cleared their school dues. The

headteacher never completed this task. Instead, the scorecard committee decided to charge a number of influential parents with encouraging other parents to pay school fees. Then class teachers distributed general circulars with information on school fees and required scholastic materials. A continual challenge is that some parents still refuse to pay. The school plans to keep on sensitizing parents and distributing circulars.

The work of many schools was similarly on-going, but in some cases, scorecard committees conducted more discrete, one-off interventions. For example, one of the schools that selected child labor as an issue for their scorecard is located on a sugar cane plantation. Many children live with their families in barracks for sugar cane laborers. A few members of the scorecard committee met with the managers of the plantation to discourage them from hiring or paying children to cut sugar cane or otherwise assist in the harvest.

The second primary data source of the process evaluation was the school rating. At two points during implementation, Elevate worked to define a strong school, an average school, and a weak school in relation to implementation of the scorecard, and to categorize each school as one of the three. The first part of this exercise generated the theory and criteria Elevate believed defined strong program implementation. Elevate identified five dimensions of program implementation: leadership, active committee, plans, attitude, and community engagement and described a vision for what strong implementation looked like for each dimension.

Just before the endline was conducted, Elevate had ranked 19 schools as strong, 22 as average, and 9 as weak. This information has been used to hypothesize the key factors for successful implementation, and the conditions or types of schools in which the scorecard can be successful. For example, does the scorecard program need a competent and motivated headteacher in order to be successfully implemented? Does the scorecard program work equally well in large and small schools?

Impact Assessment.

As treatment assignment was not random for two of the three pairs of matched subcounties, I assess the balance between treatment and control schools with data from the baseline school survey and Uwezo instrument. I report balance tables for various school characteristics (Table 1) and children's Uwezo scores (Table 2). Unsurprisingly, given that treatment assignment was completed at the subcounty level among only six subcounties, the sample was not perfectly balanced on all indicators across treatment and control. There are small differences in children's Uwezo math scores and schools' distance to a tarmacked road and PLE scores[4]. In spite of the challenges presented by the research design, the imbalance is not so great to prevent us from drawing some conclusions about the efficacy of Elevate's implementation of the scorecard program.

[4] "PLE" refers to the primary leaving exam, the standardized exam at the end of primary school. Influenced by the British colonial system, Division 1 is the best score, followed by Division 2, etc. These scores are very important within the Ugandan context. There are, however, a number of challenges in using them in analysis. Not all schools administer the PLE. Their students often sit for the exam at another nearby school. Additionally, children in private schools often have an arrangement to sit for exams at local public schools. In our endline analysis, ten schools did not administer the exam, so the sample size significantly decreases when these scores are used.

Table 1. Balance Table: School characteristics at baseline, by treatment assignment, N = 100

	Control	Treatment	Difference
School size (pupils)	359.98	344.72	15.26
	(204.08)	(165.4)	(37.15)
Fees for 3rd grade	20,212	20,760	-548
	(9,429)	(11,000)	(2,048)
Percent present in 3rd grade	78	76	2.42
	(15.67)	(19.02)	(3.48)
Km from town with tarmac road	5.265	3.936	1.329*
	(4.909)	(3.417)	(0.846)
School has a feeding program	0.60	0.58	0.02
	(0.495)	(0.499)	(0.10)
PLE pct Div 1	0.03	0.05	-0.02
	(0.064)	(0.066)	(0.014)
PLE pct Div 2	0.315	0.382	-0.067*
	(0.202)	(0.211)	(0.043)
PLE pct pass	0.604	0.701	-0.098**
	(0.260)	(0.209)	(0.050)
Number of SMC meetings	1.56	1.74	-0.18
	(0.760)	(0.853)	(0.162)
Number of PTA meetings	1.5	1.4	0.1
	(0.678)	(0.881)	(0.16)

Table 2. Balance Table: Uwezo results at baseline, by treatment assignment, N = 2,117

	Control	Treatment	Difference
English	1.993	2.039	-0.046
	(1.054)	(1.016)	(0.045)
Math	3.538	3.371	0.167***
	(1.373)	(1.369)	(0.060)

The quasi-experimental evaluation includes multiple outcomes. For parsimony, I describe one primary equation in relation to the English Uwezo scores:

$$Uwezo\ English\ Score_{ijk} = b_0 + b_1 Scorecard_k + \delta_1 X_{ijk} + \zeta_{0k} + \varepsilon_{ijk}$$

with $\zeta_{0k} \sim N(0,\sigma_\zeta^2)$ and $\varepsilon_{ik} \sim N(0,\sigma_\varepsilon^2)$ and independent of one another and where k represents the subcounty, j represents the school, and i the individual child. $Scorecard_k$ is an indicator variable for treatment assignment, X_{ijk} is a vector containing covariates.

For all outcomes, as in the primary equation, equations are multi-level mixed-effect models, accounting for the fact that children are nested in schools in subcounties. School-level outcomes, such as teacher attendance, are reported in two-level models, to account for school and subcounty clustering. Child-level outcomes, such as the English scores above, are reported in three-level models for child, school, and subcounty clustering. For dichotomous outcomes, such as transfer to school dropout, I use multi-level mixed-effects logistic regression. All analysis is Intent to Treat and children are organized by their original school, subcounty, and treatment assignment from the baseline. Tables are included at the end of the paper. Preliminary results are presented here. Future analysis will include adjustments to account for multiple hypothesis testing, missing data and robustness checks.

Transfer

In the individually-tracked student sample, 18 percent of the entire sample of children transferred to a different school between the baseline and the endline. Rates of transfer were 16% in treatment schools and 20% in control schools. The impact of the scorecard intervention on transfer is statistically significant. Children in schools with the scorecard were 0.7 times as likely to transfer as children in control schools (As a dichotomous outcome, I use multi-level mixed-effect logistic regression. Odds ratios are reported in Table 3).

This finding was supported by the results of the school survey. On average, headteachers in schools with the scorecard reported six fewer student transfers (students that left the reporting school) than headteachers in control schools (Table 4). Data from the headteacher are a two-level model, and as in the primary model, uses the general linear model.

Dropout

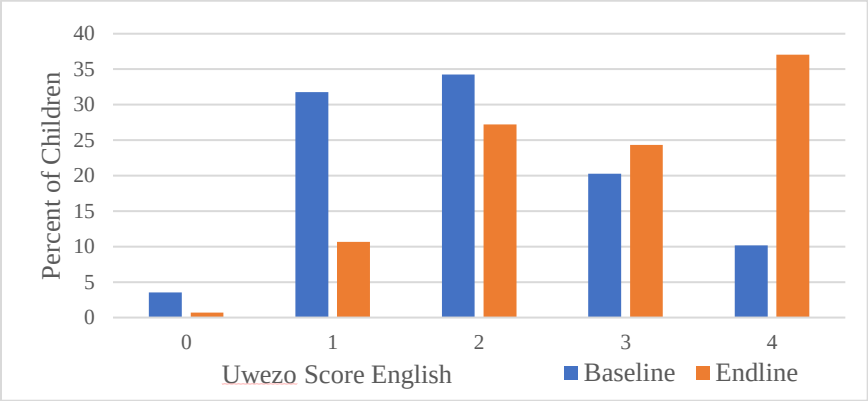
In the two years between baseline and endline, eight percent of the individually tracked student sample reported that they had dropped out of school. Dropout was also visually confirmed by the enumerators, as children were tracked to homes and workplaces in order to conduct the Uwezo assessment. Compared across treatment assignment, 9.84 percent of students in control schools and 6.24 percent of students in treatment schools reported that they had dropped out of school. As with transfer, I use a multilevel mixed-effects logistic regression to predict dropout from baseline covariates and report odds ratios in Table 5. Students in treatment schools were half as likely to dropout as students in control schools.

Seven headteachers reported that they did not know the number of children who had dropped out of their school. For the remaining 93 headteachers, on average, headteachers in treatment schools reported fewer dropouts than headteachers in control schools (3.83 as opposed to 5.27), but this difference is not statistically significant (Table 6). It is possible that limited statistical power is a factor, given the low incidence of dropout among the sample.

Endline Uwezo Scores

Between the baseline and endline, children's English and math Uwezo scores improved (English scores are illustrated in Figure 5 below). It is important to recognize, though, that children are still far below grade level. By the endline, 97% of the sample were in fourth grade or higher, but fewer than 40 percent of children could read a second-grade level story. These results are comparable to the results of Uwezo's 2015 assessment for Mukono district.

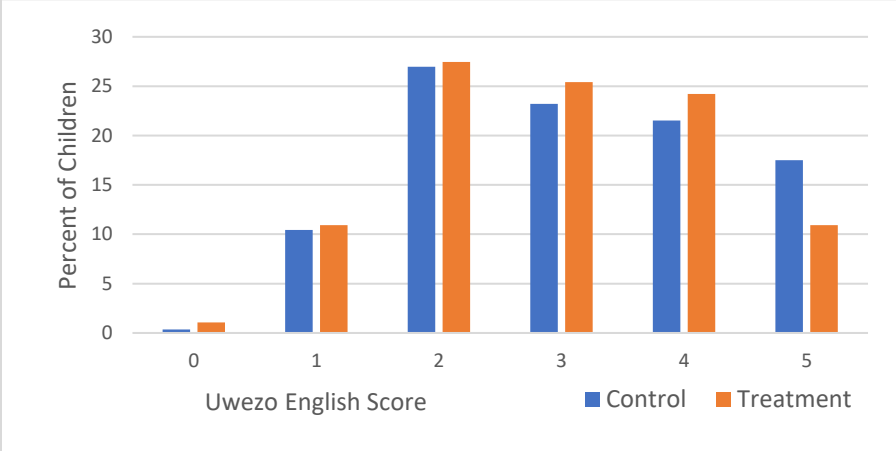
Figure 5. Baseline – Endline Comparison of Uwezo Scores



I report the results of the English analysis in Table 7. I do not find any evidence to suggest that the scorecard program impacted student's Uwezo scores when measured on a linear 0 to 5 scale[5].

It is possible that Uwezo scores do not exhibit linear properties. To explore this possibility, I compare the results of each level of the Uwezo score, illustrated in Figure 6 below. As dichotomous indicator variables, there is no difference between treatment and control for any of the levels, except for the highest, level 5, representing comprehension.

Figure 6. Endline Uwezo English Score



When dichotomized, and using multilevel mixed-effects logistic regression, it seems the participatory scorecard program decreased the likelihood that a child would reach the comprehension level (reported in Table 8 as an odds ratio less than one). While 14% of children in the control schools were at the comprehension level, only 8% of children in treatment schools were at the comprehension

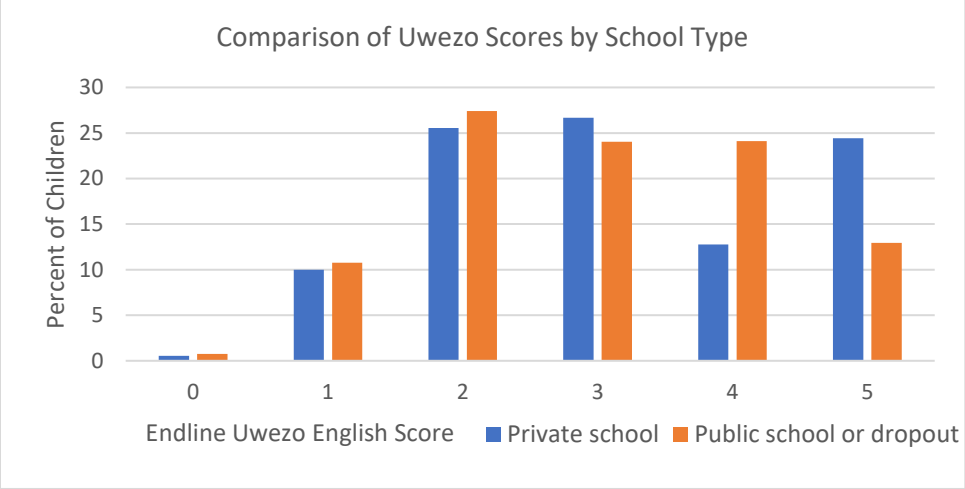
[5] As children were two years older for the endline, we included an additional level for the Uwezo scores to provide additional variation. Level four indicates that a child can read aloud a paragraph-length story while making fewer than 4 mistakes. The additional level five means that after successfully reading the paragraph, the child can correctly answer two questions about the content of the story. Uwezo refers to these questions as "bonus questions."

level. This trend was also present in the math scores, but was smaller in magnitude and was not statistically significant (Table 9).

One possible explanation for the participatory scorecard program's negative effect on student's reaching the reading level of comprehension is that children's English Uwezo scores are associated with transfer into private school. Approximately half of the children who transferred into a different school transferred into a private school. Slightly more children in control schools transferred into private schools than in treatment schools. On average, 9.47 percent of children in treatment schools transferred into a private school, while 12.27 percent of children in control schools transferred into a private school. The impact of the scorecard program on transfer into private school (Table 10) is not always statistically significant, depending upon the specification of the model, but this might be due to the limitations of the research design.

Children who transferred into a private school had slightly higher English test scores, but only in relation to the final two levels of the assessment, as illustrated in Figure 7. It is possible that in schools with the scorecard, children who would have otherwise transferred into private schools and moved from level four to level five on their Uwezo scores, stayed in their same school and did not experience the increase in their scores.

Figure 7. Comparison of endline English Uwezo scores between children who transferred into private schools and children who did not



When transfer into private school is included in the multilevel logistic regression as a covariate predicting the likelihood that a child would reach level 5 reading comprehension, children who transferred into a private school are 2.825 times more likely to score level 5 on the Uwezo English exam than children who did not transfer (Table 11).

To test whether the scorecard's impact on a child's likelihood of reaching the comprehension reading level was mediated by transfer into private school, I conduct a very basic instrumental variables analysis (Table 12). Treatment assignment's impact on transfer into private school serves as the instrument and the first stage of the two-stage least squares equation. The ultimate outcome is the dichotomous variable for the Uwezo score 5, reading comprehension, which is the second stage equation.

This analysis supports the hypothesis that the scorecard's negative impact on reading comprehension was mediated by the scorecard's negative impact on transfer into private schools. However, this analysis should be treated with caution, as the scorecard intervention may have affected many other unmeasured aspects that may predict comprehension and thus the analysis likely violates the exclusion restriction in instrumental variables analysis (Angrist, Imbens, & Rubin, 1996).

Pupil Attendance

In the data from the random spot-check, I find no impact on pupil's attendance in third grade. I also find no impact on the percent of children wearing a uniform or shoes, or with a pen, or exercise books. Across both treatment and control, I found that 82 percent of children were present. This estimate relies on a school's reported enrollment, which may be inaccurate, but I do not have any reason to believe that there would be differential reporting across treatment and control. Scorecard schools reported slightly larger P3 classes, 43.8 students as compared to 41.72, which would align with our findings on dropout, but this difference was not statistically significant. I find no impact on the pupil-teacher ratio.

Teacher Attendance

There was a significant improvement in teacher attendance across all schools between the baseline and endline. We believe this might be related to an initiative of the Mukono District Education Office. In 2017, the Officer organized a headteacher meeting in which teacher attendance was emphasized, and the DEO began conducting regular spot-checks at schools. This work was also reinforced by the District Inspector of Schools and Deputy Inspector. During the baseline, 75 percent of teachers were present at school during our spot-check. During the endline, 88 percent of teachers were present at the school.

I find a negative impact of the scorecard program on whether or not a teacher was present at the school (odds ratio 0.69) but a positive impact on whether or not the teacher was in the classroom teaching if they were at the school (odds ratio 1.27).

In Mukono, a significant portion of teachers are "PTA teachers," in that their salary is paid by the individual school with fees levied from parents. These teachers often have fewer qualifications and are hired by individual schools, rather than through the Ministry of Education. Schools with the scorecard had a slightly higher proportion of PTA teachers, 27.97 compared to 24.88, but this difference is not statistically significant.

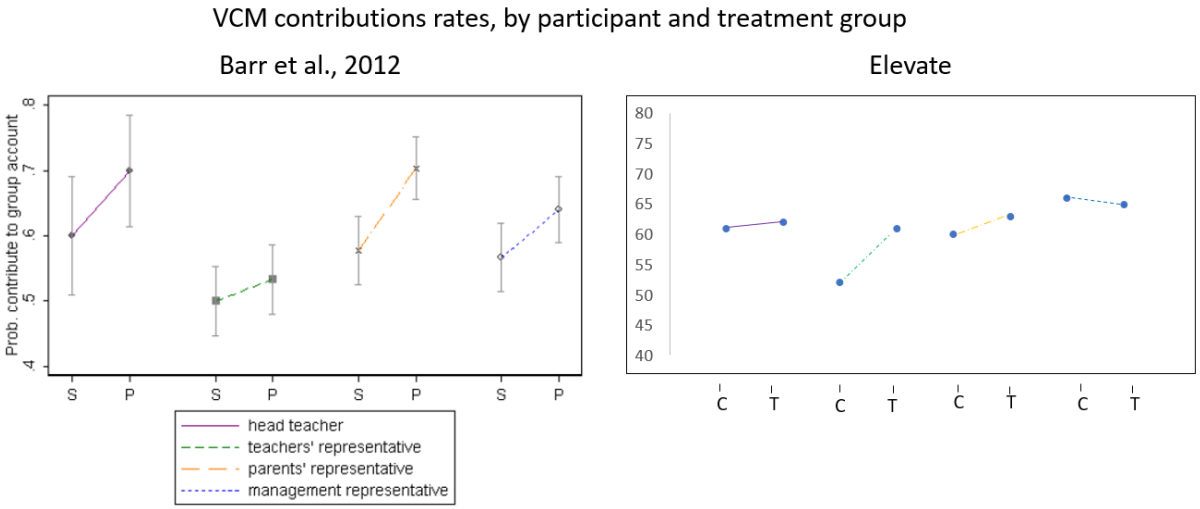
VCM Behavioral Game

Finally, I report the results of the Voluntary Contributions Mechanism (VCM) behavioral game. As in the original study, we played the game with the twelve individuals, including teachers, SMC members, the headteacher, and parents (neither study included the pupil representative). In control schools, we approximated the members of the scorecard committee as best we could. As the game was conducted during the endline, the setting and all other variables were constant across treatment and control.

I find no impact of the scorecard on the behavior of headteachers, parents, or SMC members. However, I find a positive impact of the scorecard on teachers' willingness to invest in the public good. In scorecard schools, 61 percent of teachers chose to invest in the school account. In control schools, 51 percent of teachers chose to invest in the school account. In Figure 8, I report these results alongside those of Barr et al. to show that overall, the rates are very similar (between 50 and 70 percent investment in the

school account). The Barr et al study compares the standard scorecard (S) with the participatory scorecard (P) and in our study I compare control (C) with treatment (T).

Figure 8. Comparison of VCM results between Original Study and Replication



While conducting the behavioral games, we asked all respondents a series of questions, tailored to their role (parent, SMC, teacher, headteacher). Teachers were asked five questions. Only one had a statistically significant correlation with choosing the school account. Teachers who chose the school account were much less likely to agree with the statement "Parents often blame teachers for the problems at this school." Other questions such as "I know that if I don't fulfill my duties as a teacher, there will be consequences" and "The biggest problem in primary education here in Mukono is parents' attitudes" were not correlated with the choice of account.

Preliminary Conclusions and Next Steps

Barr et al.'s (2012) study of a school scorecard program used a comparison between two different treatment variations to identify collective action as a key mechanism of impact for the one successful treatment variation. On one hand, the results of Barr et al.'s (2012) evaluation are promising: they demonstrate that a relatively inexpensive intervention can be implemented in the Ugandan context and produce positive results. On the other hand, the results also suggest a certain level of caution. The study evaluated two variations of the same intervention and only one had positive effects. Change one feature of an intervention, the study's results might imply, and it may no longer be effective.

School-based management includes a variety of different interventions and reforms. The mixed results for such initiatives might in part relate to this variety, and the vague and imprecise way in which interventions are often described. The mixed results might also be related to the work of Barr et al. (2012), and the need to differentiate between different interventions and identify how and why certain interventions work. Detailing and investigating underlying theories of change is an important direction for future research. Replicating successful interventions and understanding how they interact in different contexts, under varying conditions is another.

Understanding the mechanisms, theory of change, and enabling conditions for the participatory school scorecard program evaluated by Barr et al. (2012) is critical to determining its potential for sustainability and scale. The primary objective of this study was to investigate these concerns. Through combining qualitative investigation, process evaluation, and impact assessment, this study explores how the intervention was perceived and experienced by different stakeholders, and how general contextual factors might have translated into specific, concrete actions. Several aspects of the study are on-going, in particular, the NAPE data will provide an important link between the original study by Barr et al. (2012) and the present work.

A few working conclusions are as follows. First, the participatory scorecard intervention should be considered "balanced control" school-based management and not simply community monitoring. This is an important distinction, not just because teachers, SMC members, and the headteacher comprise the majority of the scorecard committee, and so it is an inaccurate description to describe the intervention as community monitoring, but because this categorization has implications for how the mechanisms of the intervention might be understood. There is little evidence to suggest that the scorecard works through providing parents or other stakeholders with information. There is similarly little evidence to support the theory that the scorecard works by improving accountability relationships, enabling citizens / client / community members to hold service providers accountable. Instead, it seems to be important that the scorecard encouraged communication and interaction across different stakeholdrs.

When asked how he would describe the scorecard program to someone who had never heard of it, a SMC and scorecard member responded, "It emphasizes collective effort. One voice, one voice from parents, from the school administration and even the pupils themselves. You come together and we identify some of the problems in the school." Asked the same question, a teacher and scorecard member responded "The scorecard, I think that one wants to put a relationship between the pupil, the parent, and the school." In particular, the role of parents was emphasized. Data from the process evaluation clearly illustrates that the role of parents was of primary importance. Scorecard committees' focus on parent engagement might be related to the intervention's impact on transfer and school dropout. Children remaining in the same school might be an indicator of a greater connection between the school and its community.

A second conclusion, is that the scorecard intervention should be considered within the broader education landscape and can be expected to change along with that context, whether that includes interaction with the private sector, the position of teachers, or other political or cultural factors. Within the context of Uganda, one interesting development is that SBM's focus on the school as the unit of change also places responsibility at the level of the school. Elevate's implementation of the participatory scorecard aligned with a pre-existing narrative blaming parents for the challenges facing primary education. Higher-level stakeholders such as local political leaders and officials were often aware of and involved in the intervention, but there was little discussion of their responsibility for improving primary education. In qualitative interviews, it was often recognized that parents' misconceptions surrounding education are a result of the false promises peddled by politicians, but scorecard committees focused almost exclusively on addressing parents alone.

And finally, an interesting dimension to consider in educational interventions is equity. Our data suggest that for a certain sample of students, transferring into the private sector enabled them to reach a level of reading comprehension, but they represent only 2.6 percent of all children. Of all the children who transferred into the private sector, 76 percent did not reach comprehension. One theory is that private school teachers might target children whose skills might easily be improved. Conversely, in the scorecard schools, effort seemed to focused on some of the more vulnerable students, for instance, in working to

decrease school dropout. The findings related to both transfer and dropout might indicate an increased level of support for the school. Research on SBM interventions often argues that they might take up to eight years to be fully realized (Barrera-Osorio, Patrinos, & Fasih, 2009). This study illustrates that a school scorecard program is strongly defined by the system in which it is implemented and primary education in Mukono is a complex and dynamic system in flux.

References

Angrist, J., Imbens, G., & Rubin, D. (1996). Identification of causal effects using instrumental variables. JASA Applications Invited paper, with comments and authors' response. Journal of the American Statistical Association, 91, 444-455

Banerjee, A., Banerji, R., Duflo, E., Glennerster, R., and Khemani, S. (2008), "Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education In India", NBER Working paper No. 14311.

Barr, A., Bategeka, L., Guloba, M., Kasirye, I., Mugisha, F., Serneels, P., & Zeitlin, A. (2012). Management and motivation in Ugandan primary schools: an impact evaluation report (No. 1221-2016-98593).

Barrera-Osorio, F., Patrinos, H. A., & Fasih, T. (2009). *Decentralized decision-making in schools: The theory and evidence on school-based management*. The World Bank.

Björkman, M. and Svensson, J. (2009), "Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda", Quarterly Journal of Economics 124 (2): 735.769.

Cardenas, J. and Jaramillo, C. (2007). "Cooperation in large networks: An experiment," Unpublished, Universidad de los Andes.

Cardenas, J. and Carpenter, J. "Behavioural Development Economics: Lessons from Field Labs in the Developing World," *Journal of Development Studies*, March 2008, 44 (3), 311{338.

Carr-Hill, R., Rolleston, C., Pherali, T., & Schendel, R. (2014). The effects of school-based decision making on educational outcomes in low-and middle-income contexts: A systematic review. The Campbell Collaboration.

Daily Monitor. (2011) September 5. "Government should listen to teachers" *The Daily Monitor*. *https://www.monitor.co.ug/OpEd/Editorial/689360-1230426-131crq6z/index.html*

Dauda, C. L. (2004). The importance of de facto decentralization in primary education in Sub-Saharan Africa PTAs and local accountability in Uganda. Journal of Planning Education and Research, 24(1), 28-40.

Dong, N., & Maynard, R. A. (2013). PowerUp!: A tool for calculating minimum detectable effect sizes and minimum required sample sizes for experimental and quasi-experimental design studies. Journal of Research on Educational Effectiveness, 6(1), 24-67. doi: 10.1080/19345747.2012.673143.

Forman, S. G., Shapiro, E. S., Codding, R. S., Gonzales, J. E., Reddy, L. A., Rosenfield, S. A., . . . Stoiber, K. C. (2013). Implementation science and school psychology. School Psychology Quarterly, 28(2), 77-100. http://dx.doi.org/10.1037/spq0000019

Likert, R. (1932). "A Technique for the Measurement of Attitudes," *Archives of Psychology*, 1932, Vol. 140, No. 55.

Marphatia, A. A., K. Edge, E. Legaut, and D. Archer. 2010. "Politics of Participation: Parental Support for Children's Learning and School Governance in Burundi, Malawi, Senegal and Uganda." Improving learning outcomes in Primary schools (ILOPS) Project. Research Report on Parental Participation, Action Aid, London

Malen, B., Ogawa, R.T., & Kranz, J. (1990). What do we know about school-based management ? A case study of the literature – A call for research. In W.H. Clune & J.F. Witte (Eds.) *Choice and control in American education: Vol. 2. The practice of choice, decentralization, and school restructuring.* London: Falmer.

Najjumba, I. M., Habyarimana, J & Lwanga C. (2013). Improving Learning in Uganda Vol. III: School-Based Management: Policy and Functionality. Washington, DC: World Bank. doi: 10.1596/978-0-8213-9847-0. License: Creative Commons Attribution CC BY 3.0.

Namukas, I. K., & Buye, R. (2009). Decentralisation and education in Africa: the case of Uganda. In Decentralisation, School-Based Management, and Quality (pp. 175-194). Springer, Dordrecht.

Kabay, S. (2016). Grade repetition and primary school dropout in Uganda. *Harvard Educational Review*, 86(4), 580-606.

Kabay, S. (2019). *Not always a tradeoff: Access, quality and the Global Learning Crisis.* (Unpublished Doctoral Dissertation). New York: New York University.

Kiyaga, A. and Ssenkabirwa, A. (2011) October 31. "Great expectation for ple" *The Daily Monitor*. https://www.monitor.co.ug/News/Education/688336-1264478-15disaf/index.html

Leithwood, K. and Menzies, T. (1998) "Forms and Effects of School-Based Management: A Review" *Educational Policy* Vol 12, No. 3. 325- 346.

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

Patrinos, H. A., Barrera-Osorio, F., & Fasih, T. (2009). Decentralized decision-making in schools: The theory and evidence on school-based management. The World Bank.

Rubin, H & Rubin, I. (2005) *Qualitative interviewing: The art of hearing data.* Sage Publications: Thousand Oaks, California.

Westhorp, G., Walker, B., Rogers, P., Overbeeke, N., Ball, D., & Brice, G. (2014). Enhancing community accountability, empowerment and education outcomes in low and middle-income countries: A realist review. EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

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

Table 3. Predicting Likelihood of Transfer in Individually Tracked Student Sample

	Child variables	School variables	Teacher variables	PLE
baseline English	0.920	0.989	0.986	0.937
	(0.071)	(0.083)	(0.082)	(0.082)
baseline math	1.005	0.989	0.992	1.002
	(0.055)	(0.056)	(0.056)	(0.208)
Female	1.008	0.993	1.001	0.980
	(0.138)	(0.136)	(0.136)	(0.141)
child age	0.808***	0.807***	0.808***	0.800***
	(0.045)	(0.045)	(0.045)	(0.047)
Treatment	0.719*	0.712**	0.747*	0.764
	(0.124)	(0.124)	(0.127)	(0.138)
baseline year	1.062	1.138	1.117	1.034
	(0.184)	(0.202)	(0.202)	(0.208)
km to town		1.014	1.011	1.014
		(0.020)	(0.020)	(0.021)
school size		0.999*	0.999*	0.999
		(0.020)	(0.102)	(0.001)
percent wearing shoes		1.002	1.003	1.007
		(0.007)	(0.007)	(0.008)
percent wearing uniform		1.004	1.004	1.004
		(0.004)	(0.305)	(0.004)
number of SMC meetings		1.001	0.994	1.001
		(0.106)	(0.102)	(0.109)
school average Uwezo English		0.708*	0.719	0.854
		(0.152)	(0.154)	(0.196)
school average Uwezo math		1.094	1.067	0.999
		(0.203)	(0.193)	(0.195)
teachers percent present			0.495	0.522
			(0.274)	(0.300)
teacher percent PTA			2.976*	2.363
			(1.862)	(1.604)
teacher percent teaching			1.124	1.056
			(0.506)	(0.506)
Percent pass PLE				0.434*
				(0.197)
N subcounty	6	6	6	6
N school	99	99	99	89
N students	1,657	1,657	1,657	1,491
Random-effects parameters				
sub_id	6.24e-12	3.86e-09	1.31e-06	4.35e-08
	(0.119)	(0.116)	(0.120)	(0.179)
sch_id	0.516	0.442	0.398	0.402
	(0.103)	(0.106)	(0.112)	(0.116)

Table 4. Predicting the Number of Student Transfers as Reported by Headteacher in Endline School Survey

	(1)	(2)
School size	0.004	0.002
	(0.010)	(0.012)
Km to town	0.290	0.281
	(0.324)	(0.337)
Percent wearing shoes	0.137	0.150
	(0.097)	(0.116)
Percent wearing uniform	-0.140**	-0.140**
	(0.057)	(0.060)
baseline year	-1.780	-1.941
	(2.778)	(2.980)
Treatment	-5.943**	-6.280**
	(2.641)	(2.892)
school average Uwezo English		-1.118
		(3.330)
school average Uwezo math		0.014
		(2.824)
teachers percent present		5.583
		(9.065)
teacher percent PTA		-1.041
		(10.672)
Teacher percent teaching		-1.639
		(7.104)

Figure 5. Predicting Likelihood of Dropout in Individually Tracked Student Sample

	Child variables	Teacher variables	School variables	PLE
baseline English	0.797*	0.789*	0.787*	0.797*
	(0.094)	(0.105)	(0.104)	(0.110)
baseline math	0.903	0.910	0.905	0.956
	(0.073)	(0.077)	(0.077)	(0.086)
Female	1.078	1.081	1.075	1.124
	(0.225)	(0.225)	(0.224)	(0.252)
child age	2.372***	2.364***	2.373***	2.428***
	(0.207)	(0.206)	(0.209)	(0.233)
baseline year	0.802	0.686	0.704	0.540*
	(0.200)	(0.182)	(0.197)	(0.168)
treatment	0.564**	0.538**	0.514***	0.444***
	(0.142)	(0.145)	(0.142)	(0.128)
km to town		1.002	1.004	1.001
		(0.030)	(0.030)	(0.030)
school size		0.999	0.999	0.999
		(0.001)	(0.001)	(0.001)
percent wearing shoes		1.015	1.013	1.011
		(0.010)	(0.011)	(0.011)
percent wearing uniform		1.001	1.001	1.000
		(0.006)	(0.006)	(0.006)
number of SMC meetings		0.793	0.806	0.887
		(0.124)	(0.129)	(0.151)
school average Uwezo English		0.894	0.877	0.844
		(0.282)	(0.285)	(0.286)
school average Uwezo math		0.920	0.941	0.991
		(0.253)	(0.260)	(0.284)
teachers percent present			1.854	3.023
			(1.489)	(2.532)
teacher percent PTA			0.336	0.258
			(0.332)	(0.276)
teacher percent teaching			0.856	1.180
			(0.550)	(0.800)
Percent pass PLE				0.557
				(0.376)
N sub	6	6	6	6
N sch	99	99	99	89
N stu	1,657	1,657	1,657	1,491
Random-effects Parameters				
sub_id	3.96e-07	2.24e-09	3.69e-07	1.77e-10
	(0.158)	(0.156)	(0.147)	(0.164)
sch_id	0.646	0.584	0.587	0.538
	(0.167)	(0.172)	(0.172)	(0.187)

Table 6. Predicting Number of Student Dropouts as Reported by Headteacher

	(1)	(2)
School size	0.004	0.002
	(0.004)	(0.005)
Km to town	0.141	0.132
	(0.124)	(0.124)
Percent wearing shoes	-0.036	0.004
	(0.037)	(0.043)
Percent wearing uniform	-0.029	-0.022**
	(0.022)	(0.022)
baseline year	-0.666	-0.917
	(1.158)	(1.191)
treatment	-1.011	-1.203
	(1.121)	(1.174)
school average Uwezo English		-2.447**
		(1.225)
school average Uwezo math		0.001
		(1.064)
teachers percent present		-1.224
		(3.368)
teacher percent PTA		-2.868
		(3.978)
Teacher percent teaching		1.704
		(2.687)

Table 7. Predicting Endline Uwezo Scores English

	Child variables	School variables	Teacher variables	PLE
Female	0.180***	0.184***	0.186***	0.182***
	(0.052)	(0.052)	(0.052)	(0.055)
Age	-0.153***	-0.152***	-0.152***	-0.159***
	(0.021)	(0.021)	(0.021)	(0.022)
baseline English	0.473***	0.454***	0.454***	0.448***
	(0.029)	(0.031)	(0.031)	(0.033)
baseline math	0.097***	0.097***	0.097***	0.104***
	(0.021)	(0.021)	(0.021)	(0.023)
Treatment	-0.094	-0.113	-0.090	-0.080
	(0.112)	(0.100)	(0.110)	(0.095)
km to town		-0.006	-0.006	0.000
		(0.010)	(0.011)	(0.011)
school size		0.000	0.000	-0.000
		(0.000)	0.000	(0.000)
baseline year		-0.159	-0.145	-0.141
		(0.100)	(0.110)	(0.101)
percent wearing shoes		0.003	0.003	0.007*
		(0.003)	(0.004)	(0.004)
percent wearing uniform		0.003	0.003	0.003
		(0.001)	(0.002)	(0.002)
Number of SMC meetings		-0.095*	-0.098*	-0.109*
		(0.054)	(0.055)	(0.057)
School average Uwezo English		0.113	0.128	0.100
		(0.107)	(0.111)	(0.115)
School average Uwezo math		-0.052	-0.052	-0.075
		(0.096)	(0.097)	(0.101)
Teachers percent present			-0.211	
			(0.299)	
Teacher percent PTA			0.278	
			(0.348)	
Teacher percent teaching			-0.127	
			(0.232)	
Percent pass PLE				-0.194
				(0.240)
N pupils	1,657	1,657	1,657	1,491
N schools	99	99	99	89
N subcounties	6	6	6	6
Random Effects Parameters				
Subcounty	0.0076	0.003	0.0054	1.303-16
	(0.013)	(0.011)	(0.013)	5.073-15
School	0.115	0.109	0.111	0.105
	(0.026)	(0.026)	(0.027)	(0.045)
Residual	1.008	1.007	1.007	1.040
	(0.036)	(0.036)	(0.036)	(0.050)

Table 8. Predicting Endline Uwezo English Scores: Dichotomized Level 5 Comprehension

	Child variables	School variables	Teacher variables	PLE
female	1.325*	1.347*	1.351*	1.299
	(0.231)	(0.234)	(0.235)	(0.229)
age	0.780***	0.786***	0.786***	0.779***
	(0.057)	(0.057)	(0.057)	(0.058)
baseline English	2.468***	2.394***	2.394***	2.304***
	(0.241)	(0.246)	(0.246)	(0.240)
baseline math	1.326***	1.307***	1.308***	1.313***
	(0.102)	(0.103)	(0.103)	(0.105)
treatment	0.563**	0.591**	0.608*	0.591**
	(0.151)	(0.155)	(0.608)	(0.149)
km to town		0.980	0.979	0.995
		(0.031)	(0.031)	(0.030)
school size		1.000	1.001	1.000
		(0.001)	(0.001)	(0.001)
baseline year		0.544**	0.557**	0.606*
		(0.146)	(0.155)	(0.163)
percent wearing shoes		1.010	1.008	1.018*
		(0.010)	(0.011)	(0.010)
percent wearing uniform		1.007	1.007	1.010
		(0.006)	(0.006)	(0.006)
Number of SMC meetings		0.707**	0.713**	0.630***
		(0.115)	(0.116)	(0.100)
School average Uwezo English		0.944	0.961	0.880
		(0.301)	(0.312)	(0.273)
School average Uwezo math		1.158	1.169	0.958
		(0.340)	(0.343)	(0.279)
Teachers percent present			0.808	
			(0.682)	
Teacher percent PTA			1.339	
			(1.267)	
Teacher percent teaching			0.778	
			(0.509)	
Percent pass PLE				0.956
				(0.658)
N pupils	1657	1657	1657	1491
N schools	99	99	99	89
N subcounties	6	6	6	6
Random Effects Parameters				
Subcounty	2.80e-06	1.00e-10	1.36e-10	3.19e-09
	(0.435)	(0.154)	(0.157)	(0.131)
School	0.950	0.799	0.792	0.693
	(0.143)	(0.139)	(0.140)	(0.138)

Table 9. Predicting Endline Uwezo Scores Dichotomous Level 5 Math

	Child variables	School variables	Teacher variables	PLE
Female	1.044	1.044	1.050	1.125
	(0.122)	(0.122)	(0.122)	(0.139)
Age	0.808***	0811***	0.808***	0.821***
	(0.038)	(0.038)	(0.038)	(0.041)
baseline English	1.366***	1.321***	1.316***	1.337***
	(0.094)	(0.092)	(0.091)	(0.101)
baseline math	1.216***	1.215***	1.215***	1.209***
	(0.057)	(0.058)	(0.057)	(0.062)
Treatment	0.832	0.771	0.811	0.812
	(0.169)	(0.152)	(0.156)	(0.174)
km to town		1.029	1.025	1.032
		(0.024)	(0.024)	(0.026)
school size		0.999	0.999	0.999
		(0.001)	(0.001)	(0.000)
baseline year		1.526**	1.488*	1.49
		(0.311)	(0.306)	(0.328)
percent wearing shoes		1.018**	1.016**	1.023***
		(0.008)	(0.007)	(0.008)
percent wearing uniform		1.004	1.003	1.005
		(0.004)	(0.004)	(0.005)
Number of SMC meetings		1.089	1.074	1.056
		(0.130)	(0.125)	(0.135)
School average Uwezo english		0.776		0.786
		(0.184)		(0.202)
School average Uwezo math		0.978		0.999
		(0.206)		(0.225)
Teachers percent present				
Teacher percent PTA				
Teacher percent teaching				
Percent pass PLE				0.488
				(0.265)
N pupils			1657	1491
N schools			99	99
N subcounties			6	6
Random Effects Parameters				
Subcounty			0.015	5.04e-07
			(1.26)	0.264
School			0.681	0.725
			(0.094)	(0.097)

Table 10. Predicting Transfer into Private School

	Child variables	School variables	Teacher variables	PLE
Female	0.863	0.847	0.856	0.835
	(0.144)	(0.141)	(0.143)	(0.145)
Age	0.729***	0.734***	0.735***	0.722***
	(0.050)	(0.051)	(0.051)	(0.053)
baseline English	0.936	1.056	1.049	1.014
	(0.086)	(0.107)	(0.107)	(0.107)
baseline math	0.988	0.989	0.990	1.009
	(0.065)	(0.069)	(0.068)	(0.074)
Treatment	0.719*	0.713*	0.744	0.743
	(0.139)	(0.142)	(0.145)	(0.158)
baseline year	0.863	0.903	0.859	0.896
	(0.168)	(0.184)	(0.177)	(0.247)
km to town		1.001	0.995	1.005
		(0.023)	(0.023)	(0.023)
school size		1.000	0.999	0.999
		(0.023)	(0.001)	(0.001)
percent wearing shoes		1.000	1.002	1.006
		(0.008)	(0.008)	(0.008)
percent wearing uniform		1.005	1.005	1.002
		(0.004)	(0.004)	(0.005)
Number of SMC meetings		0.987	0.983	1.028
		(0.121)	(0.115)	(0.125)
School average Uwezo english		0.581**	0.555**	0.525***
		(0.145)	(0.140)	(0.137)
School average Uwezo math		0.955	0.955	0.892
		(0.199)	(0.195)	(0.199)
Teachers percent present			0.860	0.907
			(0.544)	(0.599)
Teacher percent PTA			5.414**	6.591
			(3.850)	(5.293)
Teacher percent teaching			0.965	0.640)
			(0.504)	(0.344)
Percent pass PLE				0.855
				(0.425)
N subcounties	6	6	6	6
N schools	99	99	99	89
N students	1,657	1,657	1,657	1,491
Random Effects Parameters				
Subcounty	4.44e-09	8.40e-11	2.59e-07	0.082
	(0.117)	(0.118)	(0.169)	(0.269)
School	0.482	0.422	0.355	0.290
	(0.133)	(0.137)	(0.150)	(0.197)

Table 11. Predicting Level 5 Reading Comprehension Including Transfer into Private School

	(1)
Female	1.322
	(0.232)
Baseline English	2.516***
	(0.232)
Baseline Math	1.325***
	(0.103)
Child's age	0.804***
	(0.060)
Transfer into private school	2.825***
	(0.683)
Treatment	0.057
	(0.156)
Random-effects parameters	
Sub_id	8.92e-07
	(0.426)
Sch_id:	0.967
	(0.144)

Table 12. IV Estimates of Private School Impact on Reading Comprehension

Estimation technique	OLS	IV	IV with controls
	(1)	(2)	(3)
Private school coefficient	0.114***	2.352*	2.241*
		(1.365)	(1.206)
N students	1,657	1,657	1,657