RESULTS FOR DEVELOPMENT

Assessing Learning, Achieving Impact

A Literature Review: Citizen-led Assessments

Authors:
Molly Jamieson Eberhardt
Tara Hill
Daniel Plaut

Table of Contents

Introduction

Global efforts to reach Millennium Development Goal 2 and the Education for All Goals have resulted in a dramatic expansion of access to primary education. However, learning levels remain unacceptably low, indicating the need for a shift in focus from increasing access to increasing access to *quality* education (UNESCO 2004). As a result, arguments have been made for the inclusion of a learning goal in the renewed set of post-2015 Millennium Development Goals, which would shift the measurement of success from education outputs (such as Gross Enrollment Ratios or average years of schooling completed) to outcomes (actual assessed learning achievements), with the understanding that education for all cannot truly be achieved until the gap between enrollment and learning is closed (Filmer et al. 2006).

Central to the effectiveness of such a learning goal would be the development of a reliable, standardized, cost-effective mechanism to track learning outcomes. Unlike education access, where indicators such as school enrollment have inherent quantitative value, "there is no universal understanding of learning" (Pritchett et al. 2013) and therefore no universal system of measurement. Learning is not a clear-cut, black and white dichotomy, but instead a spectrum of skills and knowledge that hinges on a number of variables, including age, subject, context, and more. Global initiatives such as the Learning Metrics Task Force, a partnership between UNESCO and Brookings' Center for Universal Education, have sought to raise awareness of the importance of education quality in addition to access, and also to provide clarity and build consensus on how global learning indicators could be defined (UNESCO and Brookings 2013).

A wide variety of learning assessment systems have been developed and tested. One of these systems, citizen-led assessments of student learning, joins the global focus on learning with the recent rise in information-based accountability interventions designed to secure quality public service delivery for the poor (World Bank 2003). Such accountability initiatives are based on the premise that service delivery so often fails the poor because of information asymmetries between citizens, policymakers, and governments or other service providers. This is exacerbated by citizens' inability to participate or voice complaints in the service delivery process. Accountability initiatives seek to reduce such information asymmetries, providing citizens with the information and tools necessary to hold relevant stakeholders accountable, voice complaints, and monitor performance. Relatedly, citizen-led assessments aim to highlight the gap between education delivery and learning outcomes, not only bringing quality of learning to the forefront of national policy debates, but also engaging citizens in understanding the current status of education in their respective contexts, by mobilizing parents, citizens, and policymakers to take action for improved learning outcomes.

Citizen-led assessments vary according to country context with regards to design, implementation, and outcomes, but maintain several methodological principles in common. First, assessments are conducted in households in order to ensure the inclusion of out-of-school children and those attending schools outside of the mainstream public system (private, religious, and so on). Second, assessments are designed so that ordinary citizen volunteers can conduct them, measuring basic literacy and numeracy one child at a time, and using the same

basic test for all children of school age. Third, results are available instantly (meaning they can be shared with parents, if appropriate) and aggregated at district, state, and country levels, and they are disseminated to citizens, policymakers, and other education stakeholders. Finally, citizen-led assessments follow a similar theory of change across implementing organizations, namely that assessment results will increase awareness about education gaps and motivate action to improve learning outcomes at the community and/or policy level.

Results for Development Institute conducted an evaluation of the process and impact of four such citizen assessment efforts funded by the William and Flora Hewlett Foundation: ASER in India, Uwezo in East Africa, Jàngandoo in Senegal, and Beekunko in Mali. This literature review served as a foundation for the evaluation by examining existing literature on the key dimensions of effective learning assessments and other similar interventions (as well as contextual and other external factors affecting their implementation) in order to gauge the effectiveness of the citizen assessment model at measuring learning as well as inciting awareness and action to improve education policy. The first section focuses on the *design and implementation* of citizen-led assessments and the second section addresses their **attempt to foster *impact* on awareness and action.**

I. Considerations for effective assessment systems

In this section we describe the dimensions of assessment systems as derived from a survey of literature on learning assessments broadly, and discuss how each applies to the context of citizen-led assessments. We use two overarching questions to frame the seven dimensions we discuss:

> i. How should the <u>purpose</u> of an assessment influence its design?
>
> *Assessment ownership*
> *Assessment location*
> *Testing instrument*
> *Assessment frequency*
>
> ii. What considerations are necessary to ensure the <u>credibility</u> of an assessment?
>
> *Validity (testing and sampling)*
> *Reliability*
> *Language*

i. How should the *purpose* of an assessment influence its design?

The purpose of an assessment system should drive its design (Kifer 2001, USAID 2011, Wagner 2011). This section discusses four dimensions of an assessment system that must align with its intended purpose for the system to be effective: assessment ownership, assessment location, testing instrument, and assessment frequency.

Assessment ownership

While citizen-led assessments are similar to national assessments in several ways, the differences between the two are critical. The purpose of national assessments is to "describe the achievement of students in a curriculum area aggregated to provide an estimate of the achievement level in the education system as a whole at a particular age or grade level" (Greaney and Kellaghan 2008). National assessments are typically commissioned and funded (though rarely implemented) by Ministries of Education, and their audience is policymakers, media, schools, teachers, curriculum developers, and teacher trainers. Citizen-led assessments have a similar purpose, but a key difference is that they do not only intend to *describe* achievement levels but to instill in their audiences a sense of accountability so they will take action to try to raise achievement levels.

The first consideration, therefore, when it comes to who commissions and implements a citizen assessment, relates to who is holding whom accountable. One of the main audiences that citizen-led assessments aim to hold accountable is the government. As a result, it appears logical that, unlike national assessments, both the commissioner and implementer of a citizen assessment should be non-government organizations—there is a conflict of interest inherent in

the government holding itself accountable. However, this seemingly obvious separation of citizen-led assessments from government comes with a caveat.

An assessment system that aims to influence policymakers will be most successful in meeting that goal if policymakers have some level of buy-in and trust in the assessment system (Greaney and Kellaghan 2008). This is one reason to link citizen-led assessments to government structures in some—limited—way (e.g., using government colleges to source assessors as is done in some states in India or having government agencies release assessment data). This sort of link could help to soften the tension that often accompanies assessments carried out separately from the national education system (Greaney and Kellaghan 2008). On the other hand, Wagner (2011) notes that creating tension is in some cases precisely the point: "What might be called 'political defensiveness' is the other side of 'policy sensitivity,' and simply shows that measurement can be an important source of change." Nonetheless, incorporating government structures into a citizen assessment could make policymakers more likely to respond to—rather than dismiss—the assessment results.

The rationale for keeping any government engagement in citizen-led assessments limited includes the previously-mentioned conflict of interest inherent in heavily involving the government in an effort to hold its own systems accountable. Additionally, Benveniste (2002) warns that when governments run assessments, motivation can shift from measuring learning outcomes to identify areas for improvement, to measuring learning outcomes to attract investment by demonstrating adherence to international standards and donor requirements for monitoring mechanisms. This "mission drift" could result in ineffective use of assessment results.

The second consideration relating to who commissions and implements an assessment revolves around the fact that citizen-led assessments aim to motivate stakeholders *at many levels* to take action to improve learning outcomes. This includes not only policymakers but also parents, teachers, community leaders, schools, NGOs, etc. The institution responsible for the assessment must be trusted by all of these groups. Citizens at the community level (parents, neighbors, village leaders, and children) are of particular importance when it comes to considering who should commission and implement the assessment. In citizen-led assessments, a key opportunity to influence citizens at the community level is during the assessment itself, which takes place in households in order to include all children. Not all citizen-led assessments focus on sharing information (in addition to collecting it), but the opportunity exists. To capitalize on this opportunity, a citizen assessment must be driven by an institution that citizens trust to enter their villages and homes to test their children and collect personal demographic information. In some countries or areas of countries this may be a non-government organization where in others the opposite may be true. Understanding which is the case in a given context is essential to determining the appropriate roles (or combinations of roles) for government and non-government institutions. Relatedly, if the assessment is associated with a known and trusted organization, the likelihood of the acceptance and influence of results on the government and other institutions may be increased.

The third consideration relating to who commissions and implements the assessment is sustainability. Citizen-led assessments have been, to date, conducted annually. We will discuss the issue of assessment frequency below, but a large-scale assessment that repeatedly assesses a large enough sample to be representative not only at the national but sub-national level as well (as citizen-led assessments aim to do, although some first conduct pilots with smaller samples) is expensive, and may require trade-offs in terms of resource allocation that could potentially be avoided with reduced assessment frequency. These are key challenges when assessments are commissioned and implemented by independent research institutions (Greaney and Kellaghan 2008). Such institutions often have fluctuating and time-limited funding, making financial sustainability a constant variable—with associated questions about how budgeting will affect an assessment design that needs to remain consistent year after year to allow for comparability. Operational sustainability is also a challenge for large-scale assessments operating outside of an existing structure (i.e., the public education system) as thousands of individuals with varying levels of skills and experience need to be identified, trained, and deployed every time the assessment is administered.

Assessment location

Assessment administration in households as opposed to schools is a defining characteristic of citizen-led assessments. The literature cites two (non-mutually exclusive) cases in which an assessment should be administered in households as opposed to schools. The first is when the population being sampled includes all children, not only children present in school on the day of the assessment (Wagner 2011). This is central to the design of citizen-led assessments, driven by their purpose to capture learning levels of all children—including those that are not enrolled in school or are otherwise unlikely to sit for an assessment in school. As opposed to national and international assessments, which are typically designed to describe the health of an education system as defined by the performance of students enrolled, present, and assessed in a registered school, citizen-led assessments are designed to describe the health of an education system as defined by the performance of both the enfranchised and the disenfranchised—not just what the education system can do with students that actively participate, but how far it has to come to reach target learning levels for all children.

The second case in which household assessments are appropriate is when the skills and knowledge being assessed are aligned with societal expectations as opposed to the curriculum (Wagner 2011). Citizen-led assessments assess literacy and numeracy (specifically: letter and word identification, fluency in reading simple paragraphs, number recognition, single-operation computation, and, in some cases, comprehension)—which could be considered skills aligned to societal expectations depending on the context, but are more directly skills aligned to the school curriculum (albeit the same curriculum for all respondents, regardless of age). This case does not therefore drive the decision to conduct citizen-led assessments in households.

A third reason for household administration is specific to citizen-led assessments: as described above, the communities and households being assessed are one of the audiences whose action citizen-led assessments aim to stimulate (some citizen-led assessments focus on influencing this level more than others). In a household administration, parents can theoretically receive

---

immediate results, which could motivate them to put pressure on the child's teacher, school, etc. Assessment administration in the school does not allow for awareness that might lead to such action (in fact, in many assessments administered by schools the school staff does not even see the results, not to mention children or parents). The structural model of citizen-led assessments allows for this instant feedback without the allocation of significant additional resources for dissemination of findings; however there may be risks of sharing assessment results at the household level that outweigh the benefit. This tradeoff is explored as part of the evaluation.

The administration of citizen-led assessments in households is a trademark of the approach and is unquestionably aligned with the assessments' purposes for the first and third reasons described above. It is still worth noting, however, the tradeoffs involved in household assessments rather than school-based assessments. First, household assessments can be more expensive than school-based assessments (UNESCO 2008, Greaney and Kellaghan 2008, Wagner 2011) because the latter rely on an existing organizational structure (the public education system, or sometimes the education system more broadly, including non-government schools—but typically only those that are registered with the government) as well as the personnel associated with that structure as assessment administrators (Wagner 2011). Relatedly, household assessments are "difficult to conduct in comparison to in-classroom assessments," where "all the learners are grouped in one place." (Wagner 2011). Sampling validity is also relatively easier to achieve for school-based assessments, where the sampled unit (schools) can be selected at a higher level with greater quality control. In some citizen assessment models—but not all—household assessors (volunteers with limited training in the case of citizen-led assessments) are responsible for the village-level tier of the sample randomization—that is, selecting the households to be sampled—increasing the potential for criticism of the assessment's validity (this issue is discussed further below).

Household assessments of children's literacy and numeracy levels are therefore unusual (if the purpose of the assessment is to assess *adult* literacy, the assessment would need to be conducted outside of the school since adults do not attend) (Wagner 2011). The difficulty and expense of household assessments is a concession that citizen-led assessments make due to the greater relative importance of (1) sampling *all* children and (2) providing data on learning levels to the key audience of parents. The relative weight of these two issues varies among citizen assessment models.

Testing instrument

Using Clarke's framework (2011) of three types of assessment activities (classroom assessments, examinations, and large-scale assessments), citizen-led assessments can be classified as large-scale assessments. According to this definition, "large-scale assessments are designed to provide information on system performance levels and related or contributing factors (Greaney and Kellaghan 2008, Kifer 2001), typically in relation to an agreed-upon set of standards or learning goals, in order to inform educational policy and practice."

The audience for large-scale assessments includes researchers, donors, educators, national policymakers, and the general public (Beaton 1999, Greaney and Kellaghan 2008). Citizen-led assessments have the same audiences, but a critical nuance is that, in a citizen assessment, the "general public" does not only refer to an aggregate group receiving aggregate data on learning levels, but (depending on the model) also to individual parents receiving data on their child's individual learning level. As a result, the testing instrument for citizen-led assessments must produce data that is accessible to a wide spectrum of individuals, from national policymakers to often-illiterate parents. Here we discuss how the diverse audiences of citizen-led assessments necessitates a unique testing instrument that produces data that is not only immediately available and comprehensible at the household level, but also robust enough to be seen as credible at the national policy level.

To find this balance, instrument designers must ask what questions the data needs to answer in order to serve its purpose (Wagner 2011). In the case of influencing national policymakers, designers should ask, "Is it sufficient to know that there are reading problems in a population of fifth-graders in a given country or demographic group? Would the ministry like to know *why*, or when this problem began, or even *how* to begin to fix it?" (Wagner 2011). Data from citizen-led assessments can only answer the first of these questions. This is one of the tradeoffs citizen-led assessments make in order to (1) allow for rapid administration by citizen volunteers and (2) be meaningful to a diverse range of audiences. To influence the household level (parents), the instrument needs to be simple in order to produce data that lay people can understand (Wagner 2003). For both audiences, speed is also crucial: for households, due to resource constraints related to the operational difficulty of revisiting sampled households, if the data is not available to parents immediately, it will never be available; for national policymakers and other education stakeholders, if the data is not collected, analyzed, and disseminated quickly, its power to influence is limited (Wagner 2003). The rationale for this urgency is further discussed in the section below on assessment frequency.

So what should the testing instrument look like to meet the needs described here? It is useful to look at the spectrum of instruments used for other large-scale assessments. Large-scale assessments are typically broken into two categories: international and national assessments. Examples of international assessments include PISA, PIRLS, TIMSS, SACMEQ, PASEC, and Laboratorio, among others (UNESCO 2000).[1] Over 50 countries conduct national assessments (Greaney and Kellaghan 1996). Common characteristics of international assessments include a focus on assessing mastery of school curriculum organized by both content and cognitive processes (Greaney and Kellaghan 2008). For example, both PISA and TIMSS assess students' ability in mathematics content including number, geometry, algebra, data/statistics, and other areas using a combination of multiple-choice items and constructed-response items. Assessments usually take multiple days for students to complete. National assessments have a similar format.

---

[1] Programme for International Student Assessment (PISA), Progress in International Reading Literacy Study (PIRLS), Third International mathematics and Science Study (TIMSS), Southern Africa Consortium for Measuring Educational Quality (SACMEQ), Programme for the Analysis of Educational Systems of the CONFEMEN Countries (PASEC), Latin American Laboratory for Assessment of Educational Quality (Laboratorio).

On the other end of the spectrum exists a simple literacy module embedded in a household survey (such as a Demographic and Health Survey (DHS) or Living Standards Measurement Studies (LSMS)) that asks respondents to read a sentence (UN 1989, USAID 2010, Wagner 2011); the assessor then categorizes the respondent into one of two or three groups—typically some iteration of "not able to read the sentence at all, was able to read only parts of the sentence, or was able to read the whole sentence" (USAID 2010). At an even more rudimentary level, literacy modules in household surveys often ask respondents if they are able to read, without actually assessing that ability (FHI360 2008). For example, "Can (name) read personal letters, (fluently, with some difficulty or not at all) in the following languages?" (UNESCO 2008). While this would not qualify as an *assessment* as such, it remains a common approach to measuring literacy levels in various countries, and is therefore valuable when considering which testing instrument is most efficient and effective for citizen-led assessments.

Using an instrument like those sometimes embedded in household surveys ("read this sentence") may be too simple to be considered robust by national policymakers and other critical consumers of assessment data. They also lack the specificity that is often useful in making headlines—crucial for an assessment activity aiming to stimulate action. For these reasons the citizen assessment instrument must fall along the spectrum described, and not at either extreme. The Early Grade Reading Assessment (EGRA) and Early Grade Math Assessment (EGMA) developed by USAID as part of its Ed Data II initiative are examples of testing instruments in this category. They EGRA, for example, assesses students' accuracy and speed at several basic reading tasks: phonemic awareness, letter knowledge, word reading, non-word reading, oral reading fluency, and reading comprehension (USAID 2011). Each student (the assessment is administered in schools) receives a score on each task. Citizen-led assessments have a similar but simplified model that removes the element of speed (the assessment is untimed), reduces the number of tasks, and indexes each child according to the highest level task he or she was able to accurately complete.

Further discussion on the scope of the testing instrument is discussed in the section on validity below; this section has described the considerations that are involved in designing a testing instrument and assessment system that is most likely to achieve its purpose by aligning to the needs and capacities of the various audiences it aims to influence.

Assessment frequency

As with other dimensions of an assessment system's design, the frequency of the assessment should be guided by the assessment's purpose. The literature describes three recommendations for frequency of large-scale assessments based on their objectives:

Purpose of assessment results	Recommended frequency of assessment
Hold teachers, schools, students accountable for learning	Annual (Greaney and Kellaghan 2008)
Impact policy generally	Annual or biannual (Wagner 2003 and 2011)
Impact policy within annual school cycle	More than once a year (Wagner 2003)
Monitor performance of education system	Every 3-5 years (Greaney and Kellaghan 2008)

While the results of citizen-led assessments could theoretically be used for all of the purposes listed above, the links are not always direct. Citizen-led assessments may contribute to increased accountability of teachers, schools, and students—but not directly. According to the theory of change, this link would occur if education stakeholders' awareness of learning levels was increased, stimulating them to act to hold these groups accountable. As a result, it is not the assessment itself that creates the accountability—but the knowledge engendered by the assessment results. It is therefore worth considering the annual recommendation from Greaney and Kellaghan, driven by the sense of urgency that an annual assessment instills in those directly involved in the delivery of education. A major caveat is that, in the case of citizen-led assessments, schools and classes are not the unit of assessment and the same child is unlikely to be assessed two years in a row, so any motivation to perform better would be separate from the pressure of a subsequent assessment.

Of significant import to the case of citizen-led assessments are the recommendations related to policy impact. This is a major goal of citizen-led assessments. Due to the short timelines associated with policymakers' terms and their budget structures, an annual or biannual assessment allows individuals to respond to findings by initiating related policies or interventions, and still be in a decision-making role for another assessment cycle (Wagner 2003 and 2011). While policymakers may be motivated to act quickly in time to see (and take credit for) improvement in a subsequent assessment cycle, it is very unlikely that statistically significant changes would occur in one to two years, even at the lowest level where data is aggregated (Wagner 2003 and 2011).

This need for at least biannual frequency at the policy level has implications for the testing instrument as well—the data collected must be simple enough to check, aggregate, analyze, and disseminate quickly after collection (Wagner 2003). The difference between annual and biannual assessment cycles may seem minimal, but in terms of cost effectiveness it is significant. By skipping a year between assessments, training and survey administration costs could be cut nearly in half, and efforts to disseminate data to various levels could be strategically staggered to mitigate the loss of fresh data to stimulate citizen, media, and policymaker response.

The theory of change of citizen-led assessments does not include the measurement of learning outcomes over time as a means to track changing performance of the education system

(although this will inevitably happen as assessments are repeated). So, while the 3-5 year recommendation for assessments meant to monitor the system's performance makes sense given that education systems tend to change slowly (Greaney and Kellaghan 2008), the shorter time frames described above are more closely linked to citizen-led assessments' principle purposes.

That said, while measuring system performance over time is not the main purpose of citizen-led assessments, each assessment should be comparable at each of its administrations so changes can be observed over longer time periods (Greaney et al. 2009). Alterations to the assessment should therefore be minimal and procedural tools like sampling size, grading guidelines, and exclusion criteria "should be taken into account when comparisons are made between students' achievements at different points in time" (Greaney et al. 2009).

ii. What considerations are necessary to ensure the credibility of an assessment?

The rise of learning assessments as a crucial part of education policy globally has engendered a vibrant debate surrounding their technical efficacy. "Assessment in education has never been uncontroversial and it remains controversial today. Whenever an educational assessment is reported in the media, critics often challenge the results by claiming a contradictory bit of evidence, or that the assessment itself was flawed for a variety of technical reasons" (Wagner 2011). This is particularly true in the case of citizen-led assessments, which aim to hold various education stakeholders accountable; if the assessment system is not technically sound its results can easily be dismissed. Some argue that the potential policy impact of household surveys is limited, but Wagner contends that such impact is possible if the household assessment is a "serious and credible study, followed by concrete efforts to publicize results" (Wagner 2003).

Though it is inevitable that tradeoffs are made to address financial and operational constraints, citizen-led assessments need to ensure that these tradeoffs do not come at the expense of technical rigor. Here we discuss considerations involved in achieving that balance. Though most literature focused on quality standards for learning assessments has focused on classroom-based national assessments, a smaller number of studies have looked at the quality implications for household assessments of learning. Surveying both bodies of literature, this section is organized around four dimensions of an assessment system that contribute to the system's credibility, and the considerations that should be applied to each in the context of citizen-led assessments: scope and sampling (validity), reliability, and language.

Validity (testing and sampling)

The validity of an assessment system is its ability to measure what it aims to measure (Clarke 2011, Greaney and Kellaghan 2008, Wagner 2011). In the case of citizen-led assessments, validity relates to the assessment system's ability to represent children's basic literacy and numeracy levels. This can be broken down into two separate question of validity: first, do the testing instrument and the testing process adequately measure an individual child's literacy

level? Second, does the sampling methodology allow for the aggregate results to be presented as representative of the population of children in the country, the state, the district, etc.?

As previously discussed, various testing instruments are used to measure children's reading ability. Their characteristics vary depending on financial constraints, capacity of assessors, type of data needed to influence targeted audience, and population to be sampled (USAID 2011, Greaney and Kellaghan 2008). Citizen-led assessments need to produce simple but robust data in order to influence those who have a stake in the education system—from parents to policymakers. Fortunately, best practice for testing instrument design aligns with this need for simplicity: "a valid assessment instrument should measure only what it is designed to measure" (Greaney and Kellaghan 2008). It actually requires extreme care for a testing instrument to meet this seemingly basic criterion. For example, a question designed to measure a child's ability to do *subtraction* should not require the child to *read* in order to understand or begin the task. Similarly, the language and rhetoric used to explain or present assessments should not get in the way of the child accurately responding and representing his or her skill set (Wagner 2011).

While each question on the testing instrument should be very simple in order to measure only the skill it aims to, some level of nuance in assessment results beyond "literate" or "illiterate" is desirable (UN 1989): perhaps a child cannot read a paragraph but can decode words—this information is valuable for understanding how age or year in school tends to align with ability. Theoretically, testing instruments with a very large number of items aligned to various parts of the curriculum (such as the international and national assessments described above) could deliver an informative distribution of scores. However, especially when underserved households constitute a significant portion of the sampled population (as is the case with citizen-led assessments), such sophisticated testing instruments can lead to a clustering of results in the lowest performance level, which limits the usefulness of the data. This happens because international and national assessments tend to "assume children are already literate" (USAID 2011).

Citizen-led assessments fall between these two examples in terms of striking a balance between using a simple tool and producing nuanced results. Depending on the country, citizen assessment instruments have four to six task levels of increasing difficulty. A child is assigned a level based on the most difficult task he or she is able to accurately complete. For the literacy assessment, the task levels begin with letter recognition and end with reading a short story (some citizen models also include questions to test for comprehension). For the numeracy assessment, the first level is single-digit number recognition and the highest level is typically long division with a single-digit divisor. For simplicity's sake, the same instrument is used for all children, regardless of age or grade, and is typically at the first or second grade level. While this multi-tiered structure provides a lot more information than a typical literacy module in a household survey, which classifies respondents based on their ability to read one sentence, it suffers from the inverse problem that national assessments sometimes face: the data shows us that a 15-year old can read a short story, but we do not know about her ability to comprehend,

retain, or make inferences. In short, citizen-led assessments give granular results at the bottom of the skill spectrum, but leave much unknown toward the top.

Designers and implementers of citizen-led assessments put a lot of their energy into ensuring the validity of assessment results as representative of the sampled population. Wagner (2003) advises that, when it comes to sampling (or even selecting a population to study), "smaller is better": your sample should be just big enough to answer the relevant questions. Citizen-led assessments do not tend to adhere to this principle at the country level for two important reasons. First, citizen-led assessments aim to produce data that is representative not only at the national level but also at the state level and sometimes sub-state level. This means that each of those units needs to have enough data points to produce representative data across various cross-sections. For example, it is not enough to know that 60% of second-grade girls nationally cannot read a short paragraph, because each state can then claim exceptionality from the national average when confronted with the data. In order to be able to credibly claim that 60% of second-grade girls in a given state cannot read a short paragraph, a significant number of second grade girls must be assessed. When defining sampling frames, those designing citizen-led assessments should carefully consider the level at which data needs to be representative in order to trigger the desired awareness and action. The evaluation explores this issue.

The second reason is that, while technically valid, anything but a very large sample can be dismissed by policymakers as an insignificant study of only a small percentage of the population. Given that the key audiences of citizen-led assessments are not statisticians, the power of numbers is likely to be more powerful than the smallest technically-valid sample size.

Some citizen assessment models face a significant challenge with regard to their validity when it comes to the last tier of sampling. Villages are chosen for the sample by computer randomization at the national level, providing a strong foundation of sample credibility. However, in some citizen assessment models, the final tier of sampling occurs when the assessors (volunteers with limited training) arrive at their assigned villages and select the households that will be surveyed. Household selection is a systematic process that is well documented and a focus of volunteer training, but nonetheless, randomization within villages is difficult to ensure, even when volunteers have the best intentions. The degree to which the randomization of household selection is critical to the sampling methodology, which has one tier of randomization before household selection takes places, is unclear. But it nonetheless provides an opening for critics to question the validity of the assessment.

Reliability
The reliability of an assessment system is centered on the actual administration of the assessment, independent from the purpose of the assessment or the sampling methodology (these considerations were discussed previously). Reliability concerns the degree to which the assessment is capable of producing consistent and stable results (Clarke 2011); this could also be described as the degree to which "the rank ordering of individuals taking a given test would, on a second occasion, produce a very similar rank ordering" (Wagner 2011).

There are many aspects of an assessment that can influence its reliability, including the design of the instrument itself, standardization of assessment conditions, grading, length, the nature of the learner population, and other factors (Braun et al. 2006). In the case of citizen-led assessments, reliability is a critical concern because the testing tool requires the assessor to decide what level to assign a child based on his or her performance; in essence, reliability in this case is about the "repeatability of decision making" on the part of the assessor (Vagh 2012). Typically, assessments should be "administered under uniform conditions and graded according to a fixed set of rules and rubrics" for maximum reliability (Braun et al. 2006). In a household assessment, uniform conditions are unlikely—some children will be tested with neighbors and family crowding around to watch, others alone; some with limited light as the sun goes down; some inside, others outside; etc. Assessors are trained to limit distraction in the immediate area of the vicinity but in practice it is difficult to have the ideal conditions of a quiet, well-lit space without distractions or a large audience.

In regard to "rules and rubrics," citizen-led assessments do have the potential for high reliability due to the simplicity and clarity of the testing instrument and the guidelines associated with assigning children literacy and numeracy levels. Training and practice with the testing tool is a major focus of assessor training, including review of the importance of allowing children more than one attempt (the goal is to see if the child can successfully complete the task, not whether he or she can do so on the first attempt). It is crucial that such guidelines are unambiguous, with clear instructions on how many and what kind of mistakes are acceptable in the reading assessment. While these guidelines are clearly written in the assessor's manual, they may not necessarily be followed consistently in practice. The simplicity of the testing instrument itself mitigates this concern to some extent; it is easy for the assessor (often with no previous assessment or survey experience and only two days of training) to understand and requires little explanation to assessed children. That said, assessors have limited training and practice conducting the assessment under the guidance of a trainer, so it is unlikely that all rules are followed consistently (especially with regard to more nuanced rules such as what counts as a mistake in reading, how many attempts are allowed, etc.). Initial observations of assessment administration indicate that these and other areas of assessor inconsistency are quite common.

The effect of these small inconsistencies in administration on the learning level assigned to a child is unclear, but it may be quite limited. A study of two samples (n=540 and n=590) of children assessed using the ASER assessment instrument had promising results. A test-retest correlation of "substantial reliability" (0.95 in reading and 0.90 in math) was found (Vagh 2012). Additionally, the average inter-rater reliability was found to be "substantial" using Cohen's kappa coefficient, which measures the likelihood above chance (0.0) that a different assessor would give the same score to the same child (0.82 for reading and 0.79 for math, with 1.0 representing perfect agreement).

Unlike a multiple choice test, the nature of the instruments used for citizen-led assessments means that the reliability of the assessment rests on the shoulders of the assessors.[2] For this reason, the greatest investment designers and implementers of citizen-led assessments can make in the assessment's reliability is in their assessors. Greaney and Kellaghan (2008) recommend that test administrators "should have good organizational skills, have experience of working in schools, and be committed to following test and questionnaire guidelines precisely. Ideally, they should have classroom experience, speak in the same language and accent as the students, and have an authoritative but nonthreatening manner." In the case of citizen-led assessments, finding enough volunteers in every sampled district to conduct the assessment is a challenge in itself; not all of these criteria can be met. As a result, greater emphasis is placed on the *training* of volunteer assessors and their trainers. The citizen assessment model does, however, assign great importance to having assessors who speak the language of the villages they will assess. Some models are increasingly partnering with teacher education programs, increasing the likelihood that the criteria described above are met.

Language

Ensuring comparability of assessment results is a main consideration of many large-scale assessments; this includes aligning questions and indicators with global standards and establishing consistency in implementation (Wagner 2011). Citizen-led assessments are not designed to be comparable across countries, but intra-national comparison is a common feature of data dissemination efforts. As a result, designers of citizen-led assessments need to seriously consider the question of language and build mechanisms into the assessment system that allow for "flexibility and inclusion of the likely diverse population being assessed" (Wagner 2011). This ties back to the question of validity discussed previously; the assessment's validity will be compromised if it does not have a robust system for including children who speak, read, and learn in various languages (Clarke 2011).

Language can pose significant challenges to assessment designers seeking to effectively capture learning outcomes throughout an entire country. There is significant risk that children who speak minority languages will be excluded from assessments (Clarke 2011) or have their abilities misrepresented (Greaney and Kellaghan 2008). Evidence from Uganda shows that children who attend school with a language of instruction other than their mother tongue often score poorly on reading assessments, even though their ability is quite high in their mother tongue (Greaney and Kellaghan 2008). In the case of citizen-led assessments, assessing children in their mother tongue, language of instruction (and having clear guidelines about when each is done), or both is a design question with implications for validity and therefore potential impact.

A common (and understandable) response to the challenge of language is the translation of assessments into several languages. Even when assessments are translated to accommodate diverse groups of learners, "analysis must take into account the possibility that differences

---

[2] As mentioned previously, this is not an indication that multiple choice tests would be more effective; they are generally "less useful in non-school settings where low literate children may be unfamiliar with test-taking skills" (Chowdhury et al. 1994).

…may be attributable to language-related differences in the difficulty of assessment tasks," making comparability difficult (Greaney and Kellaghan 2008). Examples include alphabets with varying levels of complexity or numbers of letters and sounds (Wagner 2011). Extreme care should be taken to ensure that translation efforts take into account difficulty of tasks and use of vocabulary of equivalent familiarity to various populations (Greaney and Kellaghan 2008). For citizen-led assessments, being able to assess children throughout the country is crucial and therefore the need for testing tools in various languages is clear. It is important to keep in mind, however, the concerns about comparability of test results among various languages when findings are being packaged and disseminated.

II. How does information lead to action?

This literature review began by identifying some of the key considerations for designing effective assessment systems—the tools and processes that need to be in place to collect effective information on learning outcomes. This section will now turn to an analysis of an assessment's potential *impact,* which may or may not occur when information is leveraged in order to promote citizen, government, and/or service provider action with the ultimate goal of improving learning outcomes. The pathways that citizen action and government action take can be very different in practice. While citizen-led assessments' theories of change often stress the role that citizen action plays in generating impact (e.g., citizens receive information about low learning outcomes, feel compelled to act on the issue, take action, and government or service providers listen and respond), it is important to note that impact and action can also occur at the government or service-provider level in the absence of citizen action (e.g., a district official may hear that his or her district is performing poorly compared to others, and act without pressure from below). Existing literature and formal evaluation on the impact of information on awareness and action is more developed at the grassroots citizen level than at the policymaker level. Thus while it is important to recognize that there are various pathways through which information can lead to action and improved outcomes, this literature review will focus predominantly on the factors that can facilitate or inhibit the pathway between information and *citizen* action.

More specifically, we will provide an overview of the broader movement of information-led empowerment of citizens as a means to improve service delivery through accountability, as well as the way in which information can lead to accountability in the education sector specifically. We will explore the potential for information interventions—in particular those that take the form of learning assessments—to generate citizen action and policy reform. Importantly, this section will pay significant attention to the many factors that can influence the generation of citizen awareness and action through information, including the political and social context (the role of which is being given increased attention in the Transparency and Accountability sector – see O'Meally 2013) in which the assessment takes place, and the approach taken to the communication and dissemination of evidence. Finally, we will conclude with some of the limitations of assessments as information-for-accountability interventions, paying particular attention to their limited ability to bring about long-term improvements in learning outcomes by themselves.

i. Information, citizen activism, and government accountability

Information-for-accountability interventions that aim to improve the quality of public service provision by increasing information transparency, and by extension, encouraging citizen activism, are not limited to the education sector. Rather, academics and practitioners alike have increasingly stressed the role of citizen participation in holding governments accountable for the provision of an array of public services, including healthcare, education, water and sanitation, and electricity (World Bank 2003). The same notion of citizen-led accountability can

likewise be applied to improving the quality of private service provision. Citizens' limited knowledge of how governments and service providers perform, and their lack of information on the means through which they can engender change, play a key role in inhibiting their ability to actively hold relevant stakeholders accountable for quality service provision (Lieberman et al. 2012). Some have gone as far as to argue that "perhaps the most powerful means of increasing the voice of poor citizens in policymaking is better information" (World Bank 2003). As a result, there has been much enthusiasm surrounding interventions that seek to reduce such information asymmetries, particularly in the health and education sectors (Banerjee et al. 2010).

However, evaluations of information-based interventions show that they have achieved mixed results. Several interventions have been found to have a positive impact. For example, Pandey et al. (2007) conclude that information campaigns in a number of village clusters in Uttar Pradesh, India did improve the delivery of health and social services among poorer households, with improvements occurring in prenatal services, infant vaccinations, excess school fees, and frequency of village council meetings. Duflo et al. (2012) used a randomized experiment in India to conclude that monitoring teacher attendance daily through in-classroom cameras and providing financial incentives to teachers both reduced rates of teacher absenteeism and increased student learning outcomes. A newspaper campaign in Uganda that enabled parents and schools to monitor local officials' management of an education grant program was found to increase enrollment in schools that successfully claimed a higher percentage of their entitlements as a result of the intervention, and moreover resulted in a positive, though limited, impact on student learning outcomes (Reinikka and Svensson 2011). An intervention in Uganda that introduced community-based monitoring of public primary health care providers resulted in significant improvements in utilization and health outcomes within treatment villages (Bjorkman and Svensson 2009).

While these interventions did lead to successful outcomes, it is important to note that little qualitative work was conducted in order to understand what forces were at play between each intervention and its outcome improvement. Without more detailed process tracing and qualitative analysis, it is difficult to definitively conclude which mechanisms within the context of each intervention caused the change. Other interventions were not found to have had significant impact. Olken (2007), for example, found that bottom-up monitoring through citizen participation in village meetings had a limited effect on reducing corruption in over 600 Indonesian village road projects, as compared with the impact of a top-down monitoring approach employed by government auditors. Similarly, NGO-led monitoring of government public health facilities in India proved ineffective at improving performance outcomes eighteen months into the intervention (Banerjee et al. 2008).

The evidence to date therefore suggests that while providing citizens with information in certain cases has both empowered them and increased accountability and quality of service provision, other cases of information provision have resulted in little to no impact. These divergent findings highlight the need to better understand interventions' context as well as the ways in which citizens are provided the opportunity to participate in service delivery oversight,

both of which are aspects that play a critical role in determining the impact of any such information-for-accountability intervention (Banerjee et al. 2010).

ii. Information to action at the policymaker level

As the above examples illustrate, much of the evaluation literature on information and accountability has focused on the pathway from information to citizen action. As the evaluation field has shifted toward an increased emphasis on randomized controlled trials, information-induced policy changes which are generally occurring at a high enough level to render randomization challenging have received somewhat less attention. It is nonetheless important to also acknowledge the pathway from information to action at the policymaker or government decision-maker level and its potential to lead to improved outcomes.

Several case studies by the International Budget Partnership, for example, have illustrated the positive impact of information campaigns on policy action in the form of significant changes to budget allocations. In Tanzania, the advocacy work of a civil society group known as HakiElimu on policy and budgeting for education led to positive outcomes. In 2011, The Department of Policy and Planning decided to institute an in-service training for English and Mathematics teachers in response to the advocacy campaigns of HakiElimu, and the organization's campaign is also credited with increasing the disbursement of grants to secondary schools for teaching and learning (International Budget Partnership 2013). HakiElimu's work, while it did involve public awareness campaigns, focused in particular on directly targeting policymakers through training local government leaders, legislators, and journalists, as well as engaging in high-level lobbying and advocacy. While these cases often included some form of intermediary organization (civil society organization or think tank) leading the advocacy campaigns targeting governments, they nonetheless suggest that increased information can have a direct impact on policymakers' decisions, regardless of whether large-scale collective citizen action at the grassroots level takes place.

In fact, even in cases touted as key examples of citizen action bringing about increased accountability and policy change, scholarly debate exists as to whether such outcomes were truly the result of citizen action or not. One such example is the previously-mentioned study conducted by Reinikka and Sevensson (2011) of a public expenditure tracking survey implemented in Uganda that allowed for schools and parents to monitor local officials' handling of a large school-grant program through a newspaper-led information dissemination campaign. The results of the study found that the information campaign reduced the capture of public funds from 80 percent to under 20 percent over the course of six years. While often cited as an example of the benefits of information on citizens' ability to hold public officials accountable, some argue that these outcomes were less the result of citizen action than of policymakers' direct reaction to the information. In this particular case, Paul Hubbard argues that a previous survey conducted by the World Bank that revealed high levels of funding leakages is in fact what "strengthened the resolve within the Government of Uganda for reform and also hardened the resolve of the donor community to reduce leakages. The information campaigns

aimed at Uganda's citizens later became an element of this story, but was not the driving force" (Hubbard 2007). In terms of the type of approach that has proven most effective at influencing policymakers and governments, recently emerging school of thought in the transparency and accountability space argues that strategic engagement with government, and particularly with champions within the government, as opposed to (or in combination with) a whistle-blowing approach, may be most effective at achieving common goals of governments and civil society (Fox 2014).

It is therefore important to keep in mind that the pathway from information to action is not limited to the realm of citizen action. Positive outcomes can also result from policymakers directly taking on information and instituting changes themselves. Further work could and should be done to delineate the precise pathways from information to action at the policymaker level. Importantly, an emerging body of evidence demonstrates that improvements in governance and, relatedly, service delivery only happen with a combined effort of top-down and bottom-up accountability, ideally with a high amount of vertical integration (Fox 2014). Described by Jonathan Fox (2014) as strategic approaches, these types of multi-pronged initiatives have shown more promising results than tactical (information-only) approaches that rely heavily on assumptions about the role information alone can play in motivating citizen-based action to influence public sector performance. However, the subsequent sections will focus on the nature of the link between information and citizen action, given that this is where the majority of the literature has focused to date.

iii. Information and accountability in the context of education and learning outcomes

Bruns, Filmer, and Patrinos (2011) identify three key channels through which information can theoretically lead to a positive impact on learning outcomes: increasing choice, participation, and voice. First, providing parents with information on school performance can encourage them to *choose* higher-performing schools, while simultaneously fostering increased inter-school competition. Second, such information can also encourage parents', students', or other stakeholders' *participation* in monitoring school resource allocations and learning outcomes, ensuring closer coordination between parents' desired outcomes and the way in which resources are used. Third, access to credible evidence can empower parents and relevant stakeholders with the *voice* to pressure governments and education providers to improve policies and learning outcomes, and hold them accountable for doing so.

In the context of this framework, citizen-led assessments fall most closely under the third category of increasing voice—this happens through the collection and dissemination of evidence on learning outcomes. In theory, this can empower citizens to take action to improve learning outcomes and hold governments or service providers to a higher standard. However, citizen-led assessments are unique in that results are not only used to empower parents to demand change or take direct action, but can also be used to inform policy reforms directly. Pritchett et al. (2013) lay out several means by which evidence on learning outcomes can be used to inform education policies:

I. Setting goals for improvement

Assessments conducted on a repeated basis not only apply pressure for reform, but also enable policymakers to set realistic goals for improved learning outcomes based on an accurate picture of current achievement (Prichett et al. 2013). This is particularly relevant for citizen-led assessments, many of which are conducted on an annual basis. However, results from frequently-conducted assessments may also become predictable and thus, ineffective, particularly if they do not offer solutions for improvement.

II. Highlighting learning gaps

Assessment results can highlight learning gaps within countries at the state, district, or even local level. In addition, gaps in learning outcomes across gender, socio-economic, or ethnic lines can be revealed through assessments, thereby underscoring areas requiring policy attention (Pritchett et al. 2013). Citizen-led assessments conducted within households can also shed light on differential learning outcomes between students formally enrolled in school and out-of-school children.

III. Shedding light on what works

Assessment results can provide key insights as to what education policies and/or interventions have been successful or not.

IV. Informing realistic curriculum standards

Assessment results can help to align curriculum design with students' demonstrated knowledge (Pritchett et al. 2013). However, an assessment's ability to do this will vary depending on the design of the assessment instrument itself. Citizen-led assessments, for example, are often simple, standardized assessments of a certain grade-level, and may only be able to inform curriculum standards at a very basic level. Another potential downside to aligning results with curriculum design is that poor assessment results could also be used to lower standards and/or reduce the rigor of the curriculum.

By generating current and accurate evidence on learning outcomes, citizen-led assessments ultimately have the potential to improve those outcomes by: (1) empowering stakeholders with the evidence needed to take action, monitor service providers, and lobby education policymakers for change, and (2) informing the focus and content of reforms and policies themselves.

iv. Lessons from other information-for-accountability interventions in education

Information-for-accountability interventions in education in low and middle-income countries are not new, and several types of interventions outside of citizen-led assessments have sought to increase accountability through information generation and parental involvement regarding education outcomes. These include school report cards, test-based score rankings, participatory budgeting, and input-tracking interventions (Bruns et al. 2011), among others. Many of the lessons and best practices gained from this wider spectrum of information-based interventions are also highly relevant to the effectiveness of citizen-led assessments. A handful of such interventions are highlighted below, drawing out relevant comparisons to citizen-led assessments. The following four case studies were selected as a result of their direct relation to school performance and/or learning outcomes, in addition to the fact that each has undergone a randomized control trial or impact evaluation. Moreover, they illustrate a range of approaches taken to carrying out information interventions in the education sector (e.g., some provided information on learning outcomes specifically, others did not; some focused on school performance, others on student performance; and some were accompanied by capacity building interventions such as teacher training, while others were solely information interventions). Finally, the following four cases span a diverse range of geographic contexts.

Name of Program/Intervention	School Report Cards
Country	Pakistan
Overview	This intervention examined the impact of providing school and child-level learning report cards in 112 villages across three districts in Pakistan (for both public and private schools). Report cards were distributed to parents and teachers based on results from learning assessments in mathematics, English, and Urdu. Parents received two report cards. The first card included the child's individual score in each subject, in addition to the child's quintile rank across all tested students, and average scores for the child's school and village. The second card included the average scores for each school in the village, its quintile rank, and the number of students tested. Teachers received a supplementary card that included a break-down of scores by sub-topic. Cards were distributed during discussion groups that provided guidance on how to interpret the information in the cards. The primary findings of this intervention revealed that learning outcomes improved by 0.10 standard deviations and private school fees dropped by 21%. Andrabi et al. conclude that school report cards, through reducing schools' ability to operate in the context of information asymmetries, exert competitive pressure on schools to either increase their quality or reduce their price.

Relevance to citizen-led learning assessments	While similar to citizen-led assessments in that this intervention assessed and publicized students' learning outcomes, the channel through which impact was achieved was through increasing parents' *choice,* in contrast to citizen-led assessments, which largely focus on increasing parents' *voice*. The intervention successfully empowered parents with information about school performance and consequently enabled them to make informed decisions about where to enroll their children. Rather than facilitating parents' ability to advocate for reform or engage directly in improving their children's learning outcomes, this intervention relied more specifically on competitive market forces. This may indicate positive benefits of attributing accountability directly to individual schools, which citizen-led assessments do not do.

*Source: Andrabi et al. (2009) ; Bruns et al. (2011).

Name of Program/Intervention	**Early Grade Reading Assessment (EGRA) Plus**
Country	**Liberia**
Overview	This intervention administered the Early Grade Reading Assessment (EGRA) to three groups of schools through three distinct variations. In the first group, the EGRA was administered, but results from the assessment were not publicized or disseminated (this group served as the control group). In the second group of schools, referred to as the "light" intervention, results of the assessment were disseminated to parents through school report cards and student reading report cards prepared by teachers. Additionally, parents were informed that the test would be conducted again in the future. The third group of schools, known as the "full" intervention, benefited from the most involved approach of the three, with teachers undergoing an intensive training in reading instructional methodologies, in addition to the administration of EGRA and the dissemination of results present in the "light" intervention.

An evaluation conducted by Piper and Korda (2010) found that the "light" intervention (with information dissemination only and no teacher training) increased student outcomes in just 1 out of the 7 EGRA indicators. The more involved, "full" intervention that combined both the dissemination of results and the teacher training, however, improved student performance along *all* of the indicators measured. |
| **Relevance to citizen-led learning assessments** | This assessment-based intervention has important lessons that can be closely applied to citizen-led assessments. The divergent outcomes of the "light" and "full" intervention reflect the added positive impact of combining information dissemination with a capacity-building intervention such as teacher training. The takeaway for citizen-led assessments is that information dissemination alone is often not sufficient in and of itself to engender a meaningful improvement in learning outcomes. |

*Source: Piper and Korda (2010); Bruns et al. (2011).

Name of Program/Intervention	Village Information Campaign
Country	India
Overview	A village information campaign intervention was conducted in three Indian states: Uttar Pradesh, Madhya Pradesh, and Karnataka. While each state possessed some variation of school oversight committees (Village Education Committees in Uttar Pradesh, parent-teacher associations in Madhya Pradesh, and school development and monitoring committees in Karnataka), evidence pointed to limited levels of awareness and engagement with these bodies by community members. The information campaign delivered specific information tools to each village on a repeated basis, including a short film, a set of posters, wall paintings, take-home calendars, and learning assessment booklets. The overall goal of the campaign was to generate increased awareness about the roles and responsibilities of these committees, as well as the communities' duty to engage with them. Results from this intervention were very positive. Pandey et al. found that this information campaign not only increased levels of awareness of citizens' roles and responsibilities in these bodies, but also had a measurable impact on learning outcomes in all three states. Two states experienced a positive impact on reading levels (14-27%) in one of the three grades tested; the third state saw an improvement in writing outcomes in one grade (15%) and in mathematics results in the other grade. Levels of teacher effort also increased in two of the three states.
Relevance to citizen-led learning assessments	This intervention suggests the positive impact of information dissemination on learning outcomes. In contrast to citizen-led assessments however, this intervention did not publicize information on learning outcomes themselves, rather, on mechanisms through which parents could become involved in their children's education and hold schools accountable. Thus the information campaign directly armed parents with the knowledge necessary to take action and participate (an important point for citizen-led assessments to take into consideration when analyzing how best to incite citizen activism). It suggests that in this case parents did not necessarily require information on education outcomes or poor learning levels in order to take action – instead, they required direct information on *how* to take action and what steps concretely to take.

*Source: Pandey et al. (2008); Bruns et al. (2011)

Name of Program/Intervention	AGE (Apoyo a la Gestión Escolar)
Country	Mexico
Overview	AGE was first introduced to schools in rural Mexico in 1996-7. It seeks to increase parental participation in school management through funding parent associations and involving parents in the management and disbursement of school grants. Parents receive training in grant management and participatory skills with the goal of increasing their overall involvement in school activities. Through these parent associations, the aim is that parents will spend more time in schools, increase their levels of interaction with school staff and awareness of school issues, thereby becoming better empowered to hold schools accountable for matters ranging from teacher absenteeism to student learning outcomes and voice their opinions on school management and performance.

An evaluation of the AGE intervention (Gertler et al. 2008) found that AGE not only increased levels of parental participation in school activities and degree of parent communication with school staff (parents were more likely to be aware of and complain about teacher absenteeism and poor teaching), but moreover, had a positive impact on intermediate school quality indicators. The intervention was found to reduce grade repetition and grade failure by 4 to 5 percent. |
| Relevance to citizen-led learning assessments | Similar to citizen-led assessments, AGE seeks to increase parental awareness and involvement in their children's learning. Unlike citizen-led assessments however, this intervention facilitated parents' involvement in *schools* themselves. Parents were given access to schools through a direct mechanism – the parent associations – and were provided with the means (financial grants) and skills (through trainings) to directly take action. Without a direct mechanism for participation, the impact of information-for-accountability interventions may be limited. |

*Source: Gertler et al. (2008).

Taken together, these case studies illustrate that while information-for-accountability interventions in education do have significant potential to generate citizen action and improvement in learning outcomes, their success and impact to date have often been based on context-specific mechanisms included in each individual model. Two important lessons from the studies highlighted here have implications for the design and implementation of citizens assessments: (1) information dissemination alone is often insufficient to generate sustained and meaningful improvements in learning outcomes, or even in bringing about citizen action; and (2) important to the success of these interventions is the extent to which citizens have accessible and identifiable *channels* for action available to them (e.g., the parent associations in Mexico, or the school oversight committees in India). All too often, information interventions disproportionately focus on the provision of information itself, and under-emphasize the provision of targeted information on the means through which to act. These lessons provide an

initial framework for important considerations when evaluating the success of citizen-led assessments.

v. Assumptions and factors affecting success of information-for-accountability interventions
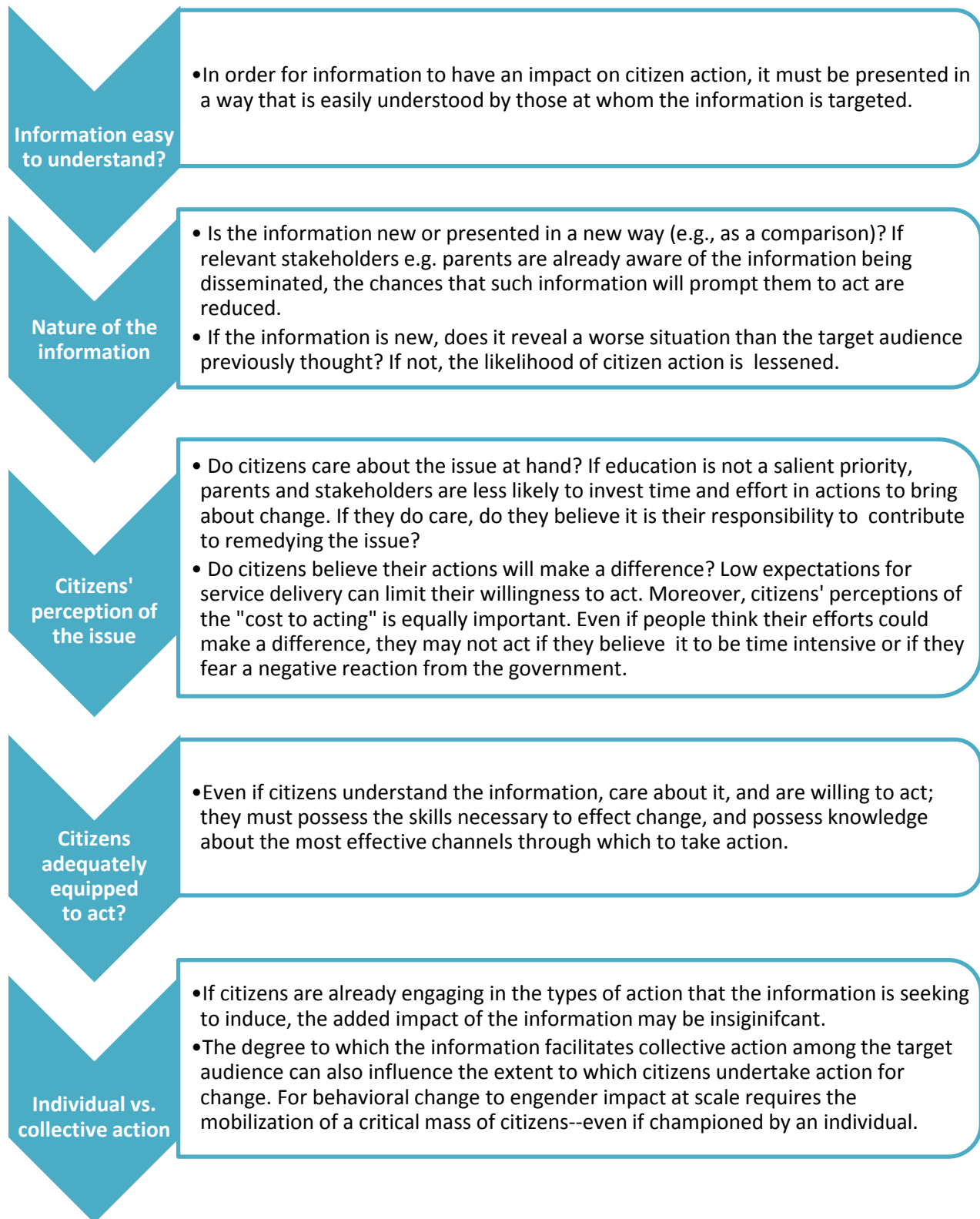
There are broadly two key considerations that affect the ability of information on learning outcomes to generate citizen action and truly hold governments to account. The first is the context and setting in which citizen-led assessments take place, and the degree to which the assumptions underlying the path from information to action are in fact present (Lieberman et al. 2012). The second is the way in which information generated from such assessments is disseminated and communicated, in addition to what information is communicated (e.g., just information about a certain school, village, or community; a cross-cutting comparison; information on what parents or other actors can do to incite changes, etc.). If certain contextual assumptions do not hold true, and if information is not disseminated in a way that encourages action, citizen-led assessments may result in limited impact.

Contextual factors

The logic underlying information-for-accountability interventions appears straightforward; increased information about underperforming service providers allows citizens to take action or apply pressure for improved outcomes, monitor performance, and advocate for policy change. In reality however, low-performing environments are often self-perpetuating and difficult to change, and exist for various context-dependent reasons (Harvard Kennedy School and Results for Development Institute 2012).

Recent research, including an evaluation of Uwezo's impact on levels of citizen action in households and villages for improved learning outcomes by Lieberman, Posner, and Tsai (2012), has underscored the role that contextual assumptions play in determining the impact of information generated from citizen-led assessments and other information-based interventions. According to Lieberman et al., the proposed link between information and government accountability induced through citizen action rests upon a key set of assumptions, outlined in what they refer to as the "Information-Accountability Causal Chain." Bruns et al. (2011) similarly identify various contextual factors that impact the ability of information to induce citizen (including parent) action. These assumptions and factors are summarized in Figure 1 on the following page.

Figure 1: Assumptions impacting the ability of information to affect awareness and action

Information easy to understand?
- In order for information to have an impact on citizen action, it must be presented in a way that is easily understood by those at whom the information is targeted.

Nature of the information
- Is the information new or presented in a new way (e.g., as a comparison)? If relevant stakeholders e.g. parents are already aware of the information being disseminated, the chances that such information will prompt them to act are reduced.
- If the information is new, does it reveal a worse situation than the target audience previously thought? If not, the likelihood of citizen action is lessened.

Citizens' perception of the issue
- Do citizens care about the issue at hand? If education is not a salient priority, parents and stakeholders are less likely to invest time and effort in actions to bring about change. If they do care, do they believe it is their responsibility to contribute to remedying the issue?
- Do citizens believe their actions will make a difference? Low expectations for service delivery can limit their willingness to act. Moreover, citizens' perceptions of the "cost to acting" is equally important. Even if people think their efforts could make a difference, they may not act if they believe it to be time intensive or if they fear a negative reaction from the government.

Citizens adequately equipped to act?
- Even if citizens understand the information, care about it, and are willing to act; they must possess the skills necessary to effect change, and possess knowledge about the most effective channels through which to take action.

Individual vs. collective action
- If citizens are already engaging in the types of action that the information is seeking to induce, the added impact of the information may be insiginifcant.
- The degree to which the information facilitates collective action among the target audience can also influence the extent to which citizens undertake action for change. For behavioral change to engender impact at scale requires the mobilization of a critical mass of citizens--even if championed by an individual.

Adapted from Lieberman et al. (2012) and Bruns et al. (2011)

It is important to note that while these assumptions and factors are presented in a simplified fashion above for the purpose of exposition, in reality they can often play out in a much more nuanced and non-linear fashion. However, what Figure 1 seeks to illustrate is that for citizen-led assessments to successfully induce awareness and action for the improvement of learning outcomes, the various contextual factors and assumptions outlined above must be taken into account. In addition to the assumptions laid out above, the responsiveness of service providers to citizen action is also a key determinant of a transparency and accountability intervention's success. Kosack and Fung (2014) outline a transparency "action cycle" comprising four main elements that must be present for a transparency policy or program to lead to improved public services: (1) information about service providers is valuable and salient to users, (2) users act on the information, (3) providers are sensitive to user actions, and (4) providers respond constructively.

A seamless pathway between information, accountability, and citizen empowerment is in reality difficult to achieve, particularly in contexts characterized by "entrenched relationships, power structures, and institutions" (Harvard Kennedy School and Results for Development Institute, 2012). The political context in which citizen-led assessments take place also influences the impact of information on levels of citizen and stakeholder action. If the nature of the government is such that citizens believe it futile or impossible to lobby for change, impact will be limited. If government policymakers do not buy into the evidence being generated by the assessments, or feel that there will be no negative repercussions for failing to act, they are unlikely to implement reform or be influenced by citizen attempts to advocate for change. For accountability to be successful, citizens must feel that the cost to monitoring and accountability is low and that its value is high; similarly, policymakers must feel that there is low cost and high value to changing their behavior (or conversely, a high cost to not changing their behavior).

Finally, in addition to the factors and assumptions outlined above, a literature review conducted of sixteen randomized control trials of various information-for-accountability interventions (Harvard Kennedy School and Results for Development Institute 2012) - many of which are referenced in this literature review - drew out several other key trends regarding the relative impact of such interventions. Notably, the review found that all of the interventions that included information on rights, or on rights and performance, were successful (success defined as having an impact on individual-level outcomes), whereas all unsuccessful interventions provided information solely on performance. Moreover, *who* delivered the information was found to be significant. All of the interventions in which citizens themselves or external researchers provided the information were successful; in cases where information was delivered by policymakers and CSOs, this was not always the case. Recent studies have shed further light on the contextual factors that influence transparency and social accountability outcomes, notably Kosack and Fung's (2014) "5 Worlds of Service Delivery" framework and O'Meally's (2013) "six contextual domains that influence social accountability."

Communication and Dissemination Approaches

Beyond the contextual factors discussed above, the process and format through which information on learning outcomes is communicated and disseminated also plays a large role in determining the impact that such information will have on levels of citizen action. Important factors that influence the effectiveness of such communication and dissemination approaches are outlined below.

I. Are the assessment results communicated in a way that is understandable and useable to all relevant stakeholders?

As previously noted, if citizens and other relevant stakeholders are unable to understand the information generated through citizen-led assessments, there is little chance that they will use the information to act. Information must be comprehensible to parents, ministry officials, and school staff for them to be able to make use of the results (Braun et al. 2006). Closely linked to this, information communication approaches should be differentiated according to the needs of distinct target audiences in order to achieve maximum impact. Tangibly, this translates into generating different forms of reports for different stakeholders, such as short overview reports that summarize key takeaways for policymakers, press releases and radio/television dissemination to reach the broader public, and specialized reports for schools, teachers, curriculum developers (Greaney and Kellaghan 2008).

Similarly, an effective strategy to communicate and disseminate assessment results should account for the fact that the various targeted stakeholders (parents, teachers, education officials, policymakers, etc.) "vary greatly in their ability to understand and apply statistical information in their decision making" (Greaney and Kellaghan 2008). Assessment results should therefore be communicated and packaged in a way that corresponds to each targeted user's level of technical expertise. One risk of information interventions that over-rely on written communication is the potential "elite capture" that may emerge as a result—whereby only literate parents are able to understand the information being disseminated (Bruns et al. 2011). This again underscores the need to tailor the information communication strategy to the needs of each audience.

II. Is the information disseminated at all levels at which citizen action and behavioral change is being sought?

The level at which information is targeted and disseminated (community, district, state, or national level) can largely depend on a citizen assessment's theory of change and intended area of impact. If the goal of a citizen assessment is to engender change at the national policy level, dissemination efforts and strategies should be focused on reaching the relevant education officers and policymakers (e.g., through tailored reports and/or meetings with relevant stakeholders). Conversely, if the goal is to provide an impetus for action at the community level through increasing parents' engagement in raising student learning outcomes, dissemination efforts must be disproportionately targeted (and tailored) to reaching audiences as the grassroots level.

Similarly, in tailoring information to each type of user at each targeted level of impact, successful communication and dissemination strategies should also consider how best to incentivize users to act. For example, passive information transmission (e.g., displaying posters) in certain information-based interventions has been found to have a limited effect on inducing parent action (Banerjee et al. 2006).

III. When and how often is information on assessment results disseminated?

The timing of information dissemination also affects the impact that assessment results can generate. Greaney and Kellaghan (2008) note that if the goal of an assessment is to affect national education policy decision-making, results should be reported as soon as possible upon the completion of the data analysis. The frequency of assessments, and by extension the frequency with which assessment results are disseminated, influences the degree of impact the results have. Disseminating data on learning outcomes each year facilitates the process of holding governments and stakeholders accountable for progress or lack thereof in learning outcomes on an annual basis. Moreover, frequent assessments and dissemination of results can contribute to an overall culture of measurement and empirical evidence-based policy making.

IV. Does the information communication strategy inform users *how* to take action?

In line with one of the key assumptions in the "Information-Accountability Causal Chain," even if assessment results are effectively communicated to the relevant stakeholders, the ultimate impact generated by the evidence will be limited unless these actors possess knowledge on *how* to leverage the information for action and change. For example, similar information-based interventions at the community and village level have revealed that even when stakeholders such as parents are interested in becoming involved in public service delivery and taking citizen action, many are "uninformed of the very institutions designed to encourage their participation" (Banerjee et al. 2006). This suggests that dissemination and communication strategies of citizen-led assessments could be made even more effective if, in addition to disseminating the results of the assessments themselves, information campaigns also communicated ways in which parents and/or other stakeholders can take action in response to the results.

vi. Citizen-led assessments alone are insufficient to impact learning outcomes

Information from citizen-led assessments can positively impact levels of citizen awareness and action, if communicated and disseminated effectively in the right context and setting. However, evidence within the literature on comparable interventions suggests that even when best practice information communication and dissemination strategies are employed, assessments alone are limited in their ability to create a long-lasting impact and improvement in learning outcomes. Assessment is not a silver bullet, and "is far from a foolproof or stand-alone remedy" (Pritchett et al. 2013). As the case studies outlined previously suggest, providing mechanisms for action alongside information provision can increase impact on stakeholder accountability and, by consequence, on learning outcomes.

Banerjee et al. (2010) analyzed the relative impact of three different types of information interventions in India that aimed to incite local participation in education outcomes. The first intervention was the most passive, simply facilitating village meetings around education through which information about local education agencies was shared with the community and village leaders. The second intervention introduced the issue of learning outcomes; members of the community were asked to gather information on enrollment and learning levels and communicate these to other community members during the village meeting. The third and most involved intervention included a specific tool to facilitate community action for improving learning outcomes. In this variation, local volunteers were trained to identify illiterate children in the community and provide them with remedial reading instruction.

The results of the study found the impact of the first two interventions to be limited. Passive information transmission had little effect on encouraging collective action. Moreover, when presented with information alone, parents' response was often to ask *what* they should do to improve learning outcomes. Introducing the tool made a difference in parents' level of engagement, who would frequently become instantly interested in the learning outcomes of their children: "Mothers would begin to push forward their children to see if they could read; when children couldn't read, there would be a sense of collective agitation and concern, and questions would begin to be raised" (Banerjee et al. 2010). In response to these findings, attention turned to the development of a specific pedagogical tool that any literate adult could use to teach children to read.

The notion that the information provision alone is not enough to bring about tangible change is supported by evidence from recent meta-analyses of transparency and accountability interventions including Fox (2014) and O'Meally (2013). This lesson likewise applies to citizen-led assessments. It suggests that the ultimate impact of citizen-led assessments on learning outcomes will be quite limited if the assessment does not have an intervention linked to it—either as part of the assessment design itself (e.g., distributing learning materials during the assessment or identifying high-scorers as potential peer leaders) or as a distinct interventional response to the assessment results. Simply making parents aware of poor learning outcomes is not enough to improve those outcomes.

vii. Limitations of assessments as a form of information-for-accountability intervention

While evidence generated from citizen-led assessments has great potential to positively impact levels of citizen engagement in holding governments and relevant stakeholders accountable for improving learning outcomes, such assessments are not without their limitations. For example, when assessment results reveal low learning outcomes, inevitably various stakeholders are held responsible (policymakers, administrators, teachers, etc.). This can result in strong opposition to reform from these stakeholders, as accepting the need for reform would be admitting their own failures and inadequacies (Bruns et al. 2011). Therefore, while citizen-led assessments may succeed in publicizing information on low learning outcomes, if those with the power to effect

change refuse to buy into the credibility of the results, overall impact measured in terms of action and reform is reduced (see Kosack and Fung 2014 for further discussion of "collaboration" vs. "confrontation" approaches in transparency and accountability interventions).

In addition, information-for-accountability interventions often lack counterfactuals, rendering it difficult to assess their value in relation to other interventions or the absence of any intervention (Lieberman et al. 2012). It is therefore important to recognize the inherent limitations of any information-for-accountability approach.

Conclusion

As the post-2015 global education agenda continues to take shape, the past decade's success in increasing access and enrollment drives our focus toward education quality and learning outcomes. In this context, the need to develop robust systems for measuring learning has become more urgent than ever. At the same time, the evidence base surrounding information-for-accountability initiatives suggests their strong potential to improve public service delivery within certain contexts. The citizen assessment model capitalizes on this unique moment, employing a hybrid approach that simultaneously measures learning outcomes and leverages information to incite citizen and policymaker action to increase accountability for quality education provision.

This literature review has surveyed the considerations that need to be made when designing an effective assessment system, as well as the channels through which citizen-led assessments have the potential to achieve impact. The first section discussed the dimensions of an assessment system that are driven by its purpose (assessment ownership, assessment location, testing instrument, and frequency), and the dimensions that determine the assessment's credibility (validity (testing and sampling), reliability, and language). Each dimension was discussed generally and applied to the specific context of citizen-led assessments.

The second section turned to a discussion of impact, providing an overview of existing information-for-accountability interventions and their relevance to citizen-led assessments. The potential of information generated by citizen-led assessments to both incite citizen action and shape education policy was highlighted. While the pathway from information to action at the policymaker level remains less developed within formal evaluation literature, it is nonetheless an important avenue for impact to keep in mind when designing information-for-accountability interventions. Emerging evidence suggests that in fact multi-pronged, top-down *and* bottom-up approaches may be most effective. Perhaps most importantly, this section provided a discussion of the contextual factors and communication and dissemination strategies that can either foster or inhibit the impact of information on various levels of citizen action, as well as the overall limitations of assessments alone to engender a lasting positive impact on learning outcomes.

This literature review served as a backdrop to the evaluation, providing a key framework for analyzing both the quality of each grantee's assessment system and the degree to which their theory of change, overall strategies, and distinct contexts have the potential to generate impact on citizen action, government accountability, and ultimately, learning outcomes.

References

Andrabi, T., Das, J., and Khwaja, A. 2009. "Report Cards: The Impact of Providing School and Child Test Scores on Educational Markets." The World Bank, Unpublished manuscript.

ASER (Rural). 2013. "Annual Status of Education Report 2012." Aser Centre.

Banerjee, A., Banerji, R., Glennerster, R., and Khemani, S. 2006. "Can Information Campaigns Spark Local Participation and Improve Outcomes? A Study of Primary Education in Uttar Pradesh, India." The World Bank Policy Research Working Paper 3967.

Banerjee, A., Banerji, R., Glennerster, R., and Khemani, S. 2010. "Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India." *American Economic Journal: Economic Policy*, 2:1.

Beaton, A. E., Postlethwaite, T. N., Ross, K. N., Spearritt, D., and Wolf, R. M. 1999. "The Benefits and Limitations of International Educational Achievement Studies." UNESCO International Institute for Educational Planning.

Benveniste, L. 2002. "The Political Structuration of Assessment: Negotiating State Power and Legitimacy." *Comparative Education Review,* 46:1.

Bourque, M. L. 2009. "A History of NAEP Achievement Levels: Issues, Implementation, and Impact 1989-2009."

Braun, H., Kanjee, A., Bettinger, E., and Kremer, M. 2006. *Improving Education Through Assessment, Innovation, and Evaluation*. American Academy of Arts and Sciences.

Bruns, B., Filmer, D., and Patrinos, H. 2011. "Making Schools Work: New Evidence on Accountability Reform." The World Bank.

Chowdhury, A. M. R., Ziegahn, L., Haque, N., Shrestha, G. L., and Ahmed, Z. 1994. "Assessing Basic Competencies: A Practice Methodology." *International Review of Education,* 40:6.

Clarke, M. 2011. "Framework for Building an Effective Student Assessment System." The World Bank READ/SABER Working Paper.

Duflo, E., Hanna, R., and Ryan, S. P. 2012. "Incentives Work: Getting Teachers to Come to School." *American Economic Review,* 102:4.

Filmer, D., Hasan, A., and Pritchett, L. 2006. "A Millennium Learning Goal: Measuring Real Progress in Education." The Center for Global Development and The World Bank Working Paper 97.

Fox, J. 2014. "Social Accountability: What Does the Evidence Really Say?" The Global Partnership for Social Accountability.

Gertler, P., Patrinos, H., and Rubio-Codina, M. 2008. "Empowering Parents to Improve Education. Evidence from Rural Mexico." The World Bank Policy Research Working Paper 3935.

Greaney, V. and Kellaghan, T. 1996. "Monitoring the Learning Outcomes of Education Systems." The World Bank.

Greaney, V. and Kellaghan, T. 2008. "Assessing National Achievement Levels in Education." The World Bank.

Greaney, V., Kellaghan, T., and Murray, T. S. 2009. "Using the Results of a National Assessment of Educational Achievement." The World Bank.

Hubbard, P. "Putting the Power of Transparency in Context: Information's Role in Reducing Corruption in Uganda's Education Sector." Center for Global Development. Working Paper Number 136.

International Budget Partnership. 2013. "Raising the Stakes: The Impact of HakiElimu's Advocacy Work on Education Policy and Budget in Tanzania."

Kifer, E. 2001. *Large-Scale Assessment: Dimensions, Dilemmas, and Policy.* Corwin Press, Inc.

Kosack, S. and Fung, A. 2014. "Does Transparency Improve Governance?" *Annual Review of Political Science*. 17:65-87.

Lieberman, E., Posner, D., and Tsai, L. 2012. "Does information lead to more active citizenship? An Evaluation of the Impact of the Uwezo Initiative in Kenya."

LMTF (Learning Metrics Task Force). 2013. "Toward Universal Learning: Recommendations from the Learning Metrics Task Force." UNESCO Institute for Statistics and Center for Universal Education at the Brookings Institution.

O'Meally, S. C. 2013. "Mapping Context for Social Accountability: A Resource Paper." Social Development Department, World Bank, Washington, DC.

Olken, B. 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy,* 115:2. The University of Chicago Press.

Pandey, P., Goyal, S., Sundararman, V. 2008. "Community Participation in Public Schools: The Impact of Information Campaigns in Three Indian States*."* The World Bank Policy Research Working Paper 4776.

Pandey, P., Sehgal, A., Riboud, M., Levine, D., and Goyal, M. 2007. "Informing Resource-Poor Populations and the Delivery of Entitled Health and Social Services in Rural India: A Cluster Randomized Controlled Trial." *The Journal of the American Medical Association*, 298:16.

Piper, B. and Korda, M. 2010. "EGRA Plus: Liberia." Program Evaluation Report, RTI International.

Pritchett, L., Banerjee, R., and Kenny, C. 2013. "Schooling is Not Education! Using Assessment to Change the Politics of Non-Learning." Center for Global Development Report.

Reinikka, R. and Svensson, J. 2011. "The Power of Information in Public Services: Evidence from Education in Uganda." *Journal of Public Economics,* 95.

UNESCO. 2000. "Education for All Status and Trends 2000: Assessing Learning Achievement." UNESCO for the International Consultative Forum on Education for All.

UNESCO. 2004. "Education for All Global Monitoring Report: The Quality Imperative." UNESCO Publishing.

UNESCO. 2008. "Using a Literacy Module in Household Surveys: A Guidebook." UNESCO Asia and Pacific Regional Bureau for Education.

United Nations. 1989. "Measuring Literacy through Household Surveys: A technical study on Literacy Assessment and Related Education Topics through Household Surveys." United Nations Department of Technical Co-operation for Development and Statistical Office.

USAID. 2010. "2010 Nigeria Education Data Survey (NEDS) Interviewer's Manual." EdData II Technical and Managerial Assistance, Task Number 8. Prepared by RTI International and National Population Commission.

USAID. 2011. "Assessing for Results: Early Grade Reading Assessments for Learning Improvement." EdData II. RTI International.

Uwezo. 2010. "Are Our Children Learning: Annual Learning Assessment Report, Uganda 2010."

Vagh, S. B. 2012. "Validating the ASER Testing Tools: Comparisons with Reading Fluency Measures and the Read India Measures." www.asercentre.org.

Wagner, D. 2003. "Smaller, Quicker, Cheaper: Alternative Strategies for Literacy Assessment in the UN Literacy Decade." *International Journal of Educational Research,* 39:3.

Wagner, D. 2011. "Smaller, Quicker, Cheaper: Improving Learning Assessments for Developing Countries." UNESCO: International Institute of Educational Planning.

World Bank. 2003. World Development Report 2004: Making Services Work for Poor People. The World Bank and The Oxford University Press.