



EdData II

Task Order 20: Measurement and Research Support to Education Strategy Goal 1

Incentives and Accountability in Education: A Literature Review

**Education Data for Decision Making (EdData II) Technical and
Managerial Assistance, Task Order 20, Activity 3
Period of Performance: 10/1/2012 – 9/30/2014
Contract Number AID-OAA-12-BC-00003
RTI Project No. 0209354.020**

June 2014, Final

This publication was produced for review by the United States Agency for International Development. It was prepared by RTI International.

Task Order 20: Measurement and Research Support to Education Strategy Goal 1

Incentives and Accountability in Education: A Literature Review

Prepared for:

Office of Education
Bureau for Economic Growth, Education, and Environment (E3)
United States Agency for International Development (USAID)
Kevin Sampson, Contracting Officer
Penelope Bender, Contracting Officer's Representative

Prepared by:

Gustavo Arcia, RTI consultant
RTI International
3040 East Cornwallis Road
Post Office Box 12194
Research Triangle Park, NC 27709-2194
Activity 3 contact: Amy Mulcahy-Dunn, amulcahy-dunn@rti.org

The author's views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development or the United States Government.

RTI International is a trade name of Research Triangle Institute.

Table of Contents

	Page
List of Figures.....	iv
List of Tables.....	iv
Abbreviations	v
Preface.....	vi
1 Introduction.....	1
1.1 Incentives or Accountability?	1
1.2 The need for a review of incentives and accountability in education	2
2 Incentives and Accountability: Concept and Approaches in School Governance.....	2
2.1 Accountability and the provision of public education.....	4
2.2 Assessment systems and test-based accountability	6
2.3 Accountability, decentralization, and school autonomy.....	6
3 Implementing Incentives and Accountability.....	8
3.1 Implementation fundamentals.....	9
3.2 Test-based accountability	10
4 Incentives and Accountability in Controlled Experiments	14
4.1 Contract teachers	15
4.2 Parent-controlled school grants	15
4.3 Performance pay for teachers.....	16
4.4 School report cards.....	17
4.5 Parent participation in school management.....	18
4.6 Accountability and education management information systems (EMIS).....	19
4.7 Education financing and system-wide incentives.....	21
4.8 Accountability and corruption in education.....	22
5 Incentives and Accountability: A Summary of Key Lessons Learned	23
5.1 Practical lessons learned for implementing incentives and accountability	23
5.2 Generalized lessons for incentives and accountability.....	25
References.....	28
Annex: Grading Formula for the Florida Comprehensive Assessment Test (FCAT).....	34

List of Figures

Figure 1. Long and short routes of accountability in public education	5
---	---

List of Tables

Table 1. Four forms of accountability in decentralized education	7
Table 2. Benefits and drawbacks of using test scores for accountability	12
Table 3. Chile's dissemination strategy for its education assessment system	19
Table 4. Example of finance-linked indicators	21

Abbreviations

E3	USAID Bureau for Economic Growth, Education, and Environment
EdData II	USAID Education Data for Decision Making project
EDUCO	Educación con Participación de la Comunidad [model, El Salvador]
EMIS	education management information system
FCAT	Florida Comprehensive Assessment Test
OECD	Organisation for Economic Co-operation and Development
PEC	Programa de Escuelas de Calidad [Quality Schools Program, Mexico]
PISA	Programme for International Student Assessment
RTI	RTI International (trade name of Research Triangle Institute)
SIMCE	Sistema de Medición de la Calidad de la Educación [Chile]
SMC	school management committee
USAID	United States Agency for International Development

Preface

To place this report into context, the United States Agency for International Development (USAID) Education Data for Decision Making (EdData II) project, led by RTI International, has a mandate to improve “*the accuracy, timeliness, and accessibility, [and use] of data for basic education policy and program planning.*” This research was carried out under EdData II Task Order 20, Measurement and Research Support to Education Goal 1, which was designed to supply a range of research and technical assistance activities to USAID/Washington to support meeting the Education Strategy’s goal of 100 million children reading by 2015. The task order has the following three objectives:

- Disseminate and extract value from EdData II data sets and experiences
- Research several aspects of effective reading improvement interventions
- Support USAID missions and host governments to set targets for reading improvement

Specifically, the study addresses a USAID request under the second objective, which was to *Conduct research on incentive and accountability programs* related to literacy instruction.

1 Introduction

The objective of this review was to learn from the successes and failures of implementing incentive and accountability mechanisms in education worldwide, to draw lessons that can be applied specifically to developing countries. The approach to the review was simple: to examine some recurrent threads or themes that suggest successful approaches and to typify to any extent possible the circumstances under which incentives and accountability succeed or fail. As such, the review begins with an examination of incentives and accountability as policy options in education, continuing later with a review of accountability and its relationship with school governance, democratic participation, and the assessment of learning.

1.1 Incentives or Accountability?

In very broad terms, an incentive is the promise of a reward for doing a good job, and accountability is the promise of a sanction if one fails to do a good job. Since the assignment of rewards implies an evaluation of the quality of the job done, accountability is embedded in the process of assigning the rewards attached to an incentive, which means that incentives and accountability are just different expressions of the same concept.

Over time, incentives in education have become a synonym for motivation driven by the promise of additional rewards for a higher level of performance, and accountability has become synonymous with the consequences of low performance. This paper discusses accountability in more detail than incentives because it assumes that producing student learning is the default option of a teaching job, and whether or not an additional incentive is promised, the evaluation of performance leading to accountability is the norm. Relatedly, incentives are discussed in terms of programs and mechanisms, such as financial rewards for teachers and students in exchange for better performance.

It is important to realize at the outset that education policies that explicitly address incentives and accountability in education are relatively recent, and that many of the existing mechanisms are yet to be evaluated for results, since they require time for implementation, time for adapting institutional structures, and time for measurements to be sufficient to sustain statistical scrutiny. Because of these constraints, the available empirical evidence directly attributable to incentives and accountability is scant. Still, there is a need to know why incentive and accountability mechanisms succeed or fail at the school level, where the pitfalls are in their implementation, and what can be said about their impact on learning outcomes.

It is also important to note that the literature on accountability—as a concept and as a program—is much more abundant than the literature on incentives, perhaps for one obvious reason: public education is paid for by public funds, and the first impulse among stakeholders is to ask for accounts, rather than to offer additional rewards. On the surface, incentives seem to be a sequel to failures of accountability within the existing system of school governance. In that vein, the use of incentives seems to be a simple attempt at correcting the failures of existing delivery systems.

1.2 The need for a review of incentives and accountability in education

Relatively recent results on controlled experiments on incentives and accountability, and on statistical analyses of international test scores within the context of school-based management, have yielded important information that can be used for designing similar programs in other countries. However, the number of these studies is relatively small. Most current efforts on incentives and accountability rely on two different approaches: participatory school-based management or trust in institutional incentives. Participatory school-based management is a practical and tested way of localizing teacher incentives and of enforcing school-level accountability in countries where institutions are far from perfect (Hanushek et al 2013). In such cases, parents use social pressure and/or delegated legal authority to ask for accounts directly from schools. The second approach, in which accountability is based on trust in the education system, is more prevalent in high-performing countries (Di Gropello 2004; Arcia, Patrinos et al. 2011), where education systems have well-established processes for rendering accounts, with resulting information that can be accessed easily by parents and by anyone else in society.

At the core of incentives and accountability in any form is the application of rewards or sanctions. Participatory school-based management allows parents to link teacher rewards or sanctions to good or bad performance. Parent-enforced accountability for teacher or school performance can be thought of as the manifestation of an educational system that does not enforce sanctions or, if it does, fails to enforce them in a fair and timely manner. In the case of trust-based accountability, systems require a functioning legal process that is charged with enforcement of fair and timely consequences.

The institutional realities of many developing countries suggest that tying a system of rewards or sanctions to accountability requires a careful alignment of personal incentives with institutional managerial incentives, with education legislation, and with everyday school management. As Crouch and Winkler (2007) point out, accountability is a necessary but insufficient condition for good governance, and implementing accountability requires paying attention to school management, teacher quality, and symmetries in information among all stakeholders.

This review is organized as follows: Section 2 reviews the concept and the definitions of accountability in education. Section 3 reviews the implementation of incentives and accountability, and section 4 reviews the different applications of accountability using controlled experiments. Section 5 draws out lessons learned and the main conclusions.

2 Incentives and Accountability: Concept and Approaches in School Governance

As indicated above, conceptually, incentives and accountability are two sides of the same coin. For purposes of this review, the term *accountability* will be used to reflect both positive and negative outcomes in school and student performance.

In general terms, accountability is a process by which a person acts in full knowledge that his or her actions imply possible positive or negative consequences (Hooge, Burns, and Wilkoszewski

2012). However, accountability can have ambiguous and multiple meanings. As argued by Bovens, Schillemans, and ‘T Hart (2008), ambiguity in its current meaning probably derives from the fact that the original Anglo-Norman concept has no direct equivalent in French, Portuguese, Spanish, German, or Dutch, where there are no semantic differences between *responsibility* and *accountability*.

In education, accountability can be interpreted as a tool for improving the effectiveness of schools and school governance, whereby stakeholders can exert direct pressure on school principals and teachers to take action to improve performance.

In practice, accountability is a process that can be structured in three phases: (1) defining results among stakeholders, (2) measuring and assessing the agreed-upon results, and (3) implementing the rewards or sanctions triggered by the results (Hooge et al. 2012). The first phase is important for those giving accounts, inasmuch as it allows them to negotiate with stakeholders what it is possible to accomplish. The second phase is important because it is a necessary component for deciding if the results merit a reward or a sanction, and the third phase is important because without rewards or sanctions, there is no accountability, just a recounting of events.

Rewards and sanctions may be substantial or minor, and the instruments used to assess accountability—such as student tests—can also be classified as high-stakes or low-stakes, depending on the size of the rewards and sanctions attached to the results. An important caveat to have in mind is that the size of the stakes in a process of accountability may be different for different actors. For example, an admission test may be of high stakes for students and of low stakes for teachers, while a test that measures student learning may be of low stakes for students but of high stakes for teachers (Rosenkvist 2010, p. 8). These occasional asymmetries suggest that implementing rewards and sanctions must be exercised with care because high-stakes accountability could have bad unintended effects.

In this regard, it must be noted that educational bureaucracies may attempt to mimic markets in assigning rewards and incentives based on objective measurements—like test scores—but using bureaucratic rules. Such managerial approaches to accountability differ significantly from other forms of accountability that are based on the perception of clients—such as parents—of what constitutes a good school performance or a good result. Noting this difference is important because accountability outside of bureaucratic rules may actually be very effective, given that it involves direct contact—and negotiation—by parents and teachers, which can produce a more complex but fine-tuned system of rewards and sanctions at the school level.¹ Still, the use of objective indicators of performance is very important for another reason: In the absence of good and pertinent information, it is common to find that parents’ perception of what constitutes a good school may differ substantially from what is considered good by education authorities. For some parents, school discipline, clean uniforms, and punctuality may be good substitutes for measures of learning. Therefore, in the long run, accountability must be placed in a context of

¹ This point was brought forward by Luis Crouch (personal communication), who also indicated that a more personalized accountability might be possible even when the client is buying from a monopoly. This effect is often found in rural areas, where teachers need to be more directly accountable to parents in order to be accepted as part of the community.

school management, which should include the regular provision by schools of pertinent information on student learning (Crouch and Winkler 2007, p. 10).

2.1 Accountability and the provision of public education

A direct approach to school accountability—especially as it relates to a government’s role in the provision of public services—is found in the World Bank’s *World Development Report 2004*, which deals with the efficiency and effectiveness of service provision for the poor (World Bank 2003). This report is important because it develops the conceptual framework for accountability in education that is currently being used by most donors, and that now is the reference point for discussing incentives and accountability in many controlled experiments around the world.

Per the World Bank guidelines, accountability in the provision of public education comprises the relationship between stakeholders and the school through five components: *Delegation, Financing, Performance, Information, and Enforcement*. As such, accountability is a relationship of power, giving stakeholders the capacity to demand that the school justify its performance, and the capacity to give rewards or impose penalties. (An expanded version of this argument can be found in Goetz and Jenkins 2002.)

Deconstructing accountability

Accountability in the provision of educational services can be described as a set of relationships among service delivery actors that has five components:

- *Delegating*. Giving explicit or implicit instructions on the services to be provided
- *Financing*. Providing or paying for the resources needed for service delivery
- *Performing*. Delivering the actual service
- *Informing*. Monitoring and evaluating performance under agreed-upon norms and standards
- *Enforcing*. Distributing rewards or sanctions tied to performance

Source: Adapted from World Bank 2003, p. 48.

In this vein, the government, as the main stakeholder in public education, *delegates* the provision of education services to schools, and supplies the necessary *financing* to pay for teachers, infrastructure, and operational expenses. Schools offer educational services under agreed-upon *performance* standards. *Information* about the assessment of school performance is used by the government to *enforce* the assignment of rewards and sanctions to education providers.

This chain of events is extremely important because the lack of any of the links in the chain reduces the possibility for accountability. This chain can follow a long bureaucratic and political route, or follow a short, direct route linking parents and schools (see Figure 1 below). *In both cases, the emerging evidence suggests two main vehicles for improving accountability in education: the implementation of assessment systems and test-based accountability, and the delegation of power to the school and to parents.*

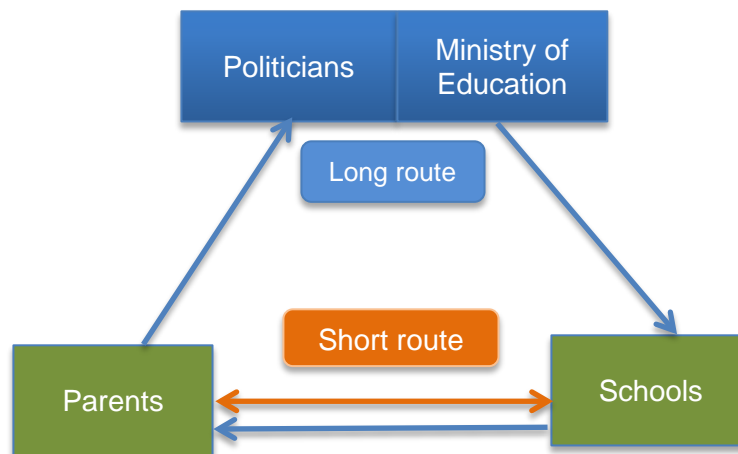
From the above process it becomes clear that in the process of accountability there is interplay among four stakeholders: (1) parents and the community, (2) politicians and policy makers, (3) local schools boards, and (4) principals and teachers. The interplay may yield accountability

relations of two forms: *weak accountability*, in which the school provides only information about results; and *strong accountability*, in which the school provides information about its performance and expects that rewards or sanctions will be forthcoming (Goetz and Jenkins 2002). One reason for this separation of two forms of accountability is that it is frequently found that the entities dealing with weak accountability tend to be different from those dealing with strong accountability. Weak accountability is generally associated with educational bureaucracies, while strong accountability is associated with school-level stakeholders.

This theoretical framework for accountability in education leads to the aforementioned *long and short routes of accountability*, in which the process of paying the consequences is filtered by the political process or implemented more directly by stakeholders. Figure 1 adapts the World Bank framework for accountability in public education. In it, parents can participate in a *long route* of accountability by expressing their concerns to politicians, who in turn use the educational structures to transform the parents' voice into action and eventual changes in education delivery. This process can take years, however, even in high-performing, mature democracies.

Figure 1. Long and short routes of accountability in public education

(Adapted from World Bank 2003)



The *short route* to accountability is one in which parents are given enough power (legal, managerial, or purchasing power) to affect school behavior directly and produce a more rapid change in education delivery. Obviously, maximal accountability is achieved when both systems are aligned.

In most school systems, it is common to find parents voicing their opinions about issues that affect their children's learning, such as the curriculum, teacher behavior, pedagogical resources, school access and discrimination, and the private cost of public education. Parents' complaints about these issues are easy to understand: They are inputs in the education process that can be easily observed, but acting on those complaints is a slow process that can be filtered by politics, by bureaucracies, and by policy designs. *The problem with the long route to accountability lies only in the inability of the system to act quickly at the school level.*

The framework for accountability shown in Figure 1 is instructive, but it needs some additional context. For incentives and accountability to work, schools must be given clear objectives and sufficient resources, and teachers must have adequate technical capabilities and personal incentives. Otherwise parental power cannot buy anything. Similarly, the expectations about the education system's performance should be aligned with the conditions on the ground; otherwise, schools and teachers will be asked for results that they cannot give, even if they want to. Schools and teachers can be made accountable if they have clear objectives and an internally coherent operational strategy with clear expectations about results, not just inputs and processes. In turn, parents should know what accounts to ask for, how to ask for accounts, and how to use accountability as a way to fix problems and motivate providers.

2.2 Assessment systems and test-based accountability

A common mechanism for assigning rewards and sanctions is the use of test scores as proximate variables for student learning. The problems of standardized testing are many, but *not* testing can be worse, since there is no way to quantify what the education system is accomplishing. This is the main reasoning behind test-based accountability (Hamilton, Stecher, and Klein 2002). On the negative side, critics argue that standardized tests measure a narrow slice of student learning; and that test-based accountability fosters distortions in school practices, such as “teaching to the test” and even cheating. In a recent blog post, Di Carlo (2012) pointed out that high school degrees are supposed to be indicators of skills that recruiters want, such as logical and critical thinking, writing ability, and math. If schools teach to a narrowly designed test that does not reflect these skills, the high school degree will eventually stop being recognized as an indicator of future success. In this case, test-based accountability will have gone too far, making test scores the end of schooling and not an indicator.

The recommendation about test-based accountability at the high school level that can be derived from the above discussion is simple: *Tests should be expanded to reflect the skills that one must have to be successful in life, and test scores should not be the prime vehicle for teacher accountability.*

In the case of early grade reading and numeracy, which are relatively narrow in scope, teaching to the test may be a perfectly reasonable strategy. However, once students acquire basic reading and numeracy, standardized tests should reflect as much as possible the skill set that would help children lead a productive life.

2.3 Accountability, decentralization, and school autonomy

Implementing incentives and accountability in decentralized education systems can be complicated but beneficial for learning. Hooge and colleagues (2012) defined four forms of school accountability in decentralized systems, of which two can be vertical—where incentives and accountability are managed by higher levels of management—and two can be horizontal—where schools manage their own system of incentives, rendering accounts to people and institutions at the same level of operational authority (see Table 1, incorporated from Hooge et al. 2012).

Table 1. Four forms of accountability in decentralized education

Vertical	<p>Regulatory school accountability: Compliance with laws and regulations; focuses on inputs and processes within the school</p> <p>Mechanism: reporting to higher levels of school authority</p>
	<p>School performance accountability: Periodic school evaluations</p> <p>Mechanisms: Standardized testing, public reporting of school performance, and rewards and sanctions</p>
Horizontal	<p>Professional school accountability: Professional standards for teachers and other educational staff</p> <p>Mechanisms: compliance with national standards for results</p>
	<p>Multiple school accountability: Continuous interactions among students, parents, and the community in assessing school management, and student and teacher performance</p> <p>Local formulation of strategies, decision-making, and evaluation, based on school-level information and local standard about results</p> <p>Mechanisms: Broader structure of incentives and sanctions</p>

Source: Adapted from Table 3.1, Hooge, Burns, and Wilkoszewski 2012.

In a highly decentralized system, vertical accountability is weak by design. For example, if power is delegated to schools and parents to design and implement their own system of rewards and sanctions, rendering accounts vertically to higher levels of government is not very useful. However, decentralized accountability must abide by national minimum standards, since autonomous schools, and by extension local stakeholders, need to follow these standards to receive funding and to evaluate school performance.

Because autonomous schools are better suited to follow the short route to accountability, the design and implementation of personal and school incentives can be highly effective. At the local level, parents can take advantage of school autonomy to establish a social contract with teachers, devising systems of incentives that are significantly broader than simple test-based accountability. This is the initial approach taken by Educación con Participación de la Comunidad (EDUCO) in El Salvador (Meza, Guzmán and de Varela 2006) and the School Autonomy program in Nicaragua (Arcia and Belli 1999). In Nicaragua, financial incentives were only a component of accountability; more parent participation in school management resulted in gains in social respect, better evaluations of performance by school directors, and more local support for the provision of inputs and materials (Arcia, Porta, and Laguna 2004).

What is interesting about a multilevel structure of accountability is the opportunity for improving school performance by combining the strength of vertical accountability (school finance, personnel management) with the strengths of horizontal accountability (broader structure of incentives, more responsive schools).

An increase in complexity in school accountability tends to create unintended effects on school performance. The pressure from stakeholders to receive clear accounts tends to produce a narrowing of teaching and learning (i.e., teaching to the test), a narrowing of the curriculum in order to focus on what is tested, and an emphasis on failure, where stakeholders associate accountability with sanctions and little else. As a bonus, high-quality teachers avoid serving low-performing students. These unintended effects increase if accountability is bureaucratic.

In contrast, accountability based on a more nuanced set of rules, applied within a context of personal negotiation—as typically happens at the school level—can be more effective in correcting simple mistakes and in motivating teachers. Test-based accountability in schools with operational autonomy should rely less on simple bureaucratic rules and more on interaction among teachers, parents, and principals, using test results as information for the negotiations (Crouch and Winkler 2007).

Horizontal accountability is well established in Europe, especially in Organisation for Economic Co-operation and Development (OECD) countries, where it is commonly associated with professional school boards, which represent the interests of parents and the community (OECD 2011). As a result, parent participation in OECD countries is not as common as in countries without professional school boards. In Europe, the empirical evidence suggests that parents tend not to participate in school affairs because professional boards do the job well, as witnessed by their high performance on Programme for International Student Assessment (PISA) tests (Arcia, Patrinos et al. 2011; Eurydice 2007). The 2012 PISA report indicated that only 11% of parents in OECD country schools participated in school government, and 79% of students reported little or no pressure from parents about the school's performance (OECD 2013). Moreover, many high-performing countries have a well-established system of school inspections—the system in the United Kingdom is a good example—that works well enough to reduce the need for parents to exercise a role of overseeing education at the school level (Jerald 2012).

Multiple school accountability is more prevalent in Latin America (Bruns, Filmer, and Patrinos 2011), but it is now being adopted by some European countries. In The Netherlands, concerns about declining test scores have led to significant increases in teacher and parent training on learning from test results (Hermans and Wiegers 2011) and in Spain, declining test scores have generated new measures for multiple school accountability (Ministerio de Educación, Cultura y Deportes 2013).

3 Implementing Incentives and Accountability

“The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to measure.”

–CAMPBELL’S LAW (as quoted by Ewing 2011)

3.1 Implementation fundamentals

Presumably, education is a desirable investment, for children; however, this investment faces a high discount rate among young people—any perceived benefits of education may be unknown or too far into the future. Intuitively, educational incentives directed to children must have a simple internal logic: the value of the incentive has to be higher than the children's discount rate. Unfortunately, this discount rate is not well understood. Hence, implementing incentives to children becomes an empirical exercise in which some incentives work and some do not. This section reviews some of the available empirical evidence.

Financial incentives to teachers are similar: why should society give additional pay to someone to do the job they were hired to do in the first place? The simplest answer is that it is easier for education systems to use money to induce changes in teaching than to reform bureaucratic procedures, education laws, or union agreements. However, as in the case of children, the discount rate of teachers is not well understood either. Vegas and Umansky (2005) argued that the structure of teacher incentives is much broader than just financial, being more of a basket of goods that includes the following nine components, of which only two are financial:

- Recognition and prestige
- Salary differentials
- Job stability
- Pension benefits
- Professional growth
- Good infrastructure and teaching materials
- A sense of mastery of teaching
- Respect from parents
- Intrinsic motivation

According to Allan and Fryer (2011), frustration with the lack of impact of educational reforms in the United States has led to experimentation with short-term financial incentives to students and teachers. These experiments are justified on grounds of cost-effectiveness: money in exchange for something may have more impact than reforms to the curriculum, reduced class size, after-school programs, teacher training, social promotion, centralized managerial control of schools, instructional coaching, increased spending, computers in the classroom, or increased parental involvement. The evidence from these experiments suggests that well-targeted incentives that pay students for reading and for completing math assignments seem to have positive significant effects on achievement. However, incentives that pay students for increased test scores do not. If students are offered money for attendance, good behavior, wearing their uniforms, and turning in their homework, students respond, but those changes do not have an impact on achievement.

The above results make sense: money in exchange for better test scores inherently assumes that students know how to be better students. This is a questionable assumption. By extension, the same argument holds for teachers: barring cheating, giving money to teachers in exchange for

higher student test scores assumes that teachers know how to be better teachers. That is another questionable assumption. Going back to the internal cost-benefit analysis done by teachers and students, the rewards for improvement have to exceed the costs inherent in increasing achievement. In the case of students, the cost of increasing test scores may be higher than the rewards, and for teachers, the cost of self-improvement to ensure better student performance may exceed the gains posed by the incentives.

Accountability is a reciprocal process. Improvements in performance should be matched with the capacity to meet the expectations (Elmore 2003). In poor countries, education systems are not generally designed or equipped to provide the support demanded by teachers and schools. The most important link between students and learning is the teacher, and teacher quality is the main obstacle to improved performance. Hence, to improve performance it is extremely important that the entire school gets focused on improving the skills teachers need to produce the learning being asked for by the school. Teacher training should be as close as possible to the place where teaching happens.

A critical problem in addressing incentives and accountability within the above framework is that education leaders and/or schools may not know what to do. Elmore argued that most low-performing schools suffer from a “lack of agreement and coherence around expectations for student learning, and they lack the means to influence instructional practice in classrooms in ways that result in student learning” (Elmore 2003, p. 9). In his view, the more prevalent the lack of expectations and lack of knowledge about classroom instruction, the lower the performance. Many low-performing schools need help setting up their own strategy for improvement and should not be penalized for low test scores during the process of doing so.

To promote better school performance, Elmore advised policy makers to develop the infrastructure for providing direct help to schools; to get the best teachers to assist low-performing schools in mentoring and training teachers on better classroom practices; and to help schools set realistic goals. External accountability should continue—so that the school remains accountable—but the criteria for accountability should match the resources available to schools so they can become accountable.

This discussion on the nature of incentives is important because it frames the use of test-based accountability as one of the main mechanisms for evaluating school performance. The other components of the basket of incentives discussed by Vegas and Umansky are largely implemented with less rigor because they are not easily quantified and monitored, and also because there is still little information on their respective impacts on learning, even though they should complement test-based criteria used for teacher evaluation (Braun 2005; Sanders et al. 2009).

3.2 Test-based accountability

In general, testing is done for the following reasons:

- To determine students’ knowledge about a particular subject
- To determine areas of strength and weakness in student knowledge

- To determine learning equity across population groups
- To learn about factors that affect student achievement
- To determine whether standards are met
- To track learning outcomes over time

This information is needed to determine incentives and rewards delivered by education authorities at the central, local, and school levels (Rosenkvist 2010). Test results promote accountability through two mechanisms: (1) their publication and (2) the use of reference standards for assigning rewards and sanctions. Under these conditions, testing may promote teaching to the test and the pursuit of a narrower curriculum (Clarke 2012).

Does testing promote learning? The evidence from the *No Child Left Behind* legislation in the United States suggests short-term gains in test scores among schools that have been penalized, but little gain among schools graded A (Rosenkvist 2010). In Chile, a test-based system of rewards for good performance showed a positive impact on schools that were close to the threshold that could move them into the top rankings, but it had little impact among schools that were already high performers or low performers (Solomon 2012).

Can students' test results be used to evaluate teachers? Learning is the result of a student's interactions with teachers, the student's own endowment, the socioeconomic conditions of students and their families, and the school environment. Isolating the effect of teachers using only test scores is very difficult. Value-added models are increasingly being used to estimate the specific contribution of individual teachers to student learning. The results are being used to determine teacher rewards and sanctions. Value-added models have two limitations: the statistical methods are very complex and prone to estimation error (Glazerman et al. 2010), and they can apply only to teachers of subjects being measured (Pearson Education 2004).

For evaluating subjects that are not assessed with standardized testing, some new metrics have proven to be very effective, such as the Early Grade Reading Assessment, which is based on a direct assessment of a student's reading ability in the first two or three years of schooling; and the Early Grade Mathematics Assessment, which does the same for math (Gove and Cvelich 2011, Reubens 2009).

Can students' test results be used to improve classroom instruction? Yes; teachers agree that test results can be useful for making pedagogical adjustments. The problem is that teachers disagree on what aspects of the results to use in making pedagogical adjustments and how to use them. This is compounded by the lack of teacher training on how to design tests or draw meaning from test data. According to Rosenkvist (2010), these problems are worst in schools that most need pedagogical adjustment. Table 2 (drawn in part from Rosenkvist 2010) summarizes the advantages and disadvantages of test-based accountability.

Table 2. Benefits and drawbacks of using test scores for accountability

Use	Benefits	Drawbacks
Generate performance tables	<ul style="list-style-type: none"> • Comparison stimulates improvement • Teachers are responsive to their schools' rating and ranking • Results involve and inform parents • Schools are held accountable 	<ul style="list-style-type: none"> • Teachers dislike school rankings • Parents do not use the full information • It is statistically difficult to distinguish differences among most schools • Schools with low capacity seldom improve
Issue rewards and sanctions	<ul style="list-style-type: none"> • Consequences can have a positive effect on student outcomes • Teachers and schools receive incentives to improve instruction 	<ul style="list-style-type: none"> • Narrowing of the curriculum often follows • More resources are allocated to subjects being tested • Results focus on students close to the threshold score • Test scores are vulnerable to manipulation • Teachers and schools can be rewarded or penalized for factors out of their control
Evaluate teachers	<ul style="list-style-type: none"> • Teachers are held accountable • Value-added assessment can isolate teacher effect • Evaluation results can support teacher certification • Assessments identify teacher training needs 	<ul style="list-style-type: none"> • Test captures only a portion of the contribution of teachers and schools • Methodological problems exist in value-added assessment • Only some subjects are assessed • Teachers lack training in designing tests and analyzing and using test data

Source: Columns 1 and 2: Adapted from Rosenkvist 2010, p. 30; column 3: author.

Can the positive effects of testing compensate the drawbacks? Based on the OECD experience, Rosenkvist (2010) concluded that schools whose personnel know how to interpret test scores and how to relate them to the factors associated with their students' learning make good decisions about teachers, support systems, and pedagogical adjustments.

Test-based accountability is politically convenient, since it sends political leaders clear indications about the effect of education policy on learning (McDonnell 2002). Having objective measurement of results helps politicians reduce the asymmetries in information, where school principals and teachers know more than politicians about what goes on inside the schools. It is precisely the problem of information asymmetries that creates conflict between politicians and education systems during the process of interpreting test results.

Because the general public tends to look at test results as good indicators of education quality, politicians are eager to use them to gather political support. For example, the State of Florida applies the Florida Comprehensive Assessment Test (FCAT) to grade public schools. Even though Florida parents are regularly reminded of the limitations and distortions of high-stakes testing, at the end of the day, parents make schooling decisions and related household decisions on the basis of the FCAT rankings (see Annex for the FCAT grading formula).²

² In Florida there is evidence that FCAT scores significantly affect housing prices. In 1999 the average price of a house near an A school was about \$9,000 higher than the price of a house near a B school (Figlio and Lucas 2000). Many real estate websites in Florida now use FCAT rankings to help potential buyers decide where to shop for a house.

West and Peterson (2006) examined the effects of the State of Florida's system for grading schools based on their FCAT performance. A year after its implementation, students in *F* schools—which faced intervention by the state—increased performance by 0.04 standard deviations higher than students in non-*F* schools. Moreover, the authors found that other types of accountability used in Florida that were not tied to sanctions had no impact on student performance. Because schools in Florida are funded mostly by property taxes, A schools tend to receive more money indirectly, since good schools tend to increase housing prices in an area. The Florida example leaves open the question whether threats work better than incentives (or vice versa) and whether there are lessons learned to scale up test-based accountability.

Aligning incentives with test-based accountability implies a clear definition of the support that will be provided to those teachers and schools with low-performing students. Typically, teachers would like more professional training, better pedagogical materials, and more flexibility in the way they conduct their classes, although there is evidence that (particularly in going from poor to fair levels of performance) less flexibility, such as the use of more scripted lessons, may be more effective (Mourshed, Chijioke, and Barber 2010).

Hout and Elliott (2011) considered that in using test-based accountability, one must take into account the need to be aware of the ways in which student testing can distort instructional behavior, such as teaching to the test or gaming the incentives. If defined too narrowly, incentives can promote distortions in teacher or student behavior. Experienced teachers can figure out ways in which incentives can be distorted or captured without doing what incentives are supposed to promote.

For example, teachers tend to pay more attention to students just under the threshold of proficiency, since that is where the performance gains are (Hout and Elliott 2011, p. 16). Similarly, teachers may allocate significant amounts of time to test-taking skills instead of teaching the subject (Koretz 2008) or help students during high-stakes testing to ensure improved scores (Kane and Staiger 2002).

These caveats lead to the following conclusion: There is a need to (1) explore other incentives aside from test-based ones, (2) ensure that the basic design of test-based incentives minimize distortions, and (3) monitor existing incentives to improve them over time.

In the case of *multiple school accountability*, there are plenty of examples where multiple stakeholders participate in school decisions, but there is little published evidence on the type of training they receive in school-based management, or in self-evaluation to determine the effect on school quality. Since multiple school accountability involves a broader, continuous interaction among the school, parents, students, and the community, most current efforts at school autonomy and decentralization down to the school level involve elements of this approach.

A typical example of multiple school accountability is the Quality Schools Program in Mexico (called PEC, the Spanish acronym for *Programa de Escuelas de Calidad*), a federal program that fosters improved learning through the *voluntary* implementation of school-based management practices. The program, begun in 2001 and now covering more than 50,000 schools, is being

expanded nationwide. Any public school in Mexico can participate in the program, but preference is given to schools in poor and disadvantaged areas.

PEC gives schools a small grant (approximately US\$4,000) to be spent on any improvements decided on in collaboration with the parents' council. The program trains principals in school management practices, including parent participation, planning and organization, and accountability; and the school is given complete autonomy over the use of grant funds and the operational budget.³

By law, all PEC schools must have an active school council. Council members receive training and manuals. Parents participate in all aspects of school governance except teacher deployments and transfers. The council must agree on the school's expenditure priorities.

A recent econometric analysis by Arcia, Patrinos, and Rivera-Olvera (2012) concluded that PEC does have a positive impact on learning outcomes, especially in mathematics at the primary level. Standardized test scores data for the past five years showed that PEC schools had average scores that were 4 percent of a standard deviation higher than non-PEC schools. The results also showed that poverty had a significant negative effect on test scores and that this effect was highest in the early grades of primary school.

Recent case study analyses of the best PEC schools showed that the program had led to high levels of satisfaction among parents and teachers in the implementation of school-based management practices (Arcia and Rivera-Olvera 2012). Moreover, the case study evidence strongly suggested that the role of the school principal and the level of implementation of PEC management practices were crucial for improving education quality.

The PEC example was just a case study; because this type of accountability is still loosely defined, there has been no systematic assessment of the impact of multiple school accountability on student outcomes.

4 Incentives and Accountability in Controlled Experiments

The main methodological problem with analyzing school incentives is that it is difficult to obtain data of sufficient quantity and quality to discern the impact of a given incentive and correctly attribute it to each of the different participants in the education process (Gertler et al. 2011). To help with this methodological problem, some investigators have started implementing controlled experiments, which would be better for determining the exact impact of rewards and sanctions on teacher and student behavior.

This section reviews different types of *controlled experiments on school incentives*. The list is not exhaustive, as there is a continuous flow of new studies in the literature, but it shows the ones

³ For a brief history of PEC and the current regulation framework for its operations, consult the government's newspaper of record, *Diario Oficial de la Federación*, at http://dof.gob.mx/nota_detalle.php?codigo=5228060&fecha=26/12/2011.

that have been reported online and thus are generally accessible. This section reports on several incentives tested in controlled experiments: contract teachers, performance pay, parent control of school grants, school report cards, and parent participation in school-based management. In addition, this section examines the available evidence on the impact of some incentives that were evaluated using quasi-experimental methods, which are not as rigorous as controlled experiments but can give better information than case studies or non-experimental designs.

4.1 Contract teachers

An experiment with contract teachers and reduced class sizes in the first grade in 210 schools in Kenya found that locally hired contract teachers had less absenteeism and that learning outcomes improved by 0.16 of a standard deviation (Duflo, Dupas, and Kremer 2012), a difference roughly equal to an additional three months of schooling (Hill et al. 2007). The authors of this study also found that local civil service teachers undermined contract teachers because the latter had much lower salaries and produced better outcomes, creating a threat to civil servants. During the experiment, civil service teachers worked less and pushed for the hiring of relatives as contract teachers.⁴

Parent council members in the experiment received 90 minutes of training on monitoring teacher performance, and additional training on the contract-teacher program, including how to supervise the hiring of a teacher, how to do an interview, how to ask other parents for input, and how to monitor teacher attendance. Parent council members were also trained in how to hold meetings, create a calendar of activities, do a final evaluation of contract teachers, and complete a performance report. The results suggest that contract teacher programs work well if school-based management is in place and if parent councils associations have some power over school decisions.

In another randomized controlled experiment on the impact of contract teachers on learning outcomes in Kenya (Bold et al. 2013), it was found that test scores went up among schools supervised by a nongovernmental organization and did not go up in schools supervised by the government. Differences in outcomes seem to have been caused by the hiring of friends and family members of civil service teachers in schools under weak government supervision, a significant problem identified by Kremer, Moulin, and Namunyu (2003).

4.2 Parent-controlled school grants

A controlled experiment on school grants was conducted in The Gambia as part of its Whole School Development program (Blimpo and Evans 2011). In this experiment, one group of schools received a grant, with school staff and parents receiving management training; another group of schools received the grant and no training; and a third group served as the control, without a grant or the training. A total of 273 schools participated in the program. Teacher absenteeism and student attendance were reduced by more than 20% in treatment schools, but test scores were the same among treatment and control schools. In villages with high initial

⁴ This finding also highlights the problems of randomized trials when double blinding is impossible: contamination, both positive and negative, is nearly impossible to prevent.

literacy, the program was very successful; but in villages with low literacy, the program had negative effects. For grant-only schools, there was no impact on student or teacher attendance.

Low parent literacy was a hurdle inasmuch it precluded parents from understanding school and student performance. Despite many efforts, school and student performance was not well explained by schools even after four years in the program. The authors cautioned against implementing school-based management programs in poor, undeveloped areas.

A pilot project in Indonesia tested four methods for empowering school management committees (SMCs): (1) giving block grants to SMCs, (2) training SMC members in school management, (3) establishing democratic elections of SMC members, and (4) creating formal linkages between the SMC and the village council (World Bank 2011).

The results showed that the provision of block grants was the most helpful tool, followed by democratic elections of SMC members. Training was effective if tied to visits to good schools by SMC members from other schools. The use of elections of SMC members was positively linked with homework supervision by parents and with time on task by teachers. Of the four interventions, the *block grants were most effective because they were used to increase spending on student activities, and to hire more temporary staff. Closer parent participation increased teacher attendance. These three actions had a significant effect on raising student scores in language and mathematics.*

4.3 Performance pay for teachers

This section reviews two selected controlled experiments with teacher pay increases tied to student performance. Glewwe, Ilias, and Kremer (2010) analyzed the impact of linking teacher salaries to student test scores in Kenya. The program was very successful after the second year, when teachers understood how it worked. Most of the *salary gains were linked to a higher number of students taking the test*; the program penalized teachers for every student that did not take the test. However, the *gains in learning outcomes happened only in the areas being tested, while learning in other parts of the curriculum did not increase*. Test score gains were higher on multiple-choice exams than on fill-in-the-blank questions, indicating that students and teachers had become adept at test taking. Absenteeism did not improve; what changed was the time given by teachers to the areas that were linked to the test.

In rural Andhra Pradesh, India, Muralidharan (2012) analyzed the impact of a five-year experiment with teacher pay linked to performance in public primary schools. Researchers tracked a cohort of primary school students for the five years of primary school. The results showed that *the teacher performance pay program had positive, significant effects on learning outcomes of the covered topics, and positive spillover effects on learning outcome on non-covered topics*. Students served by performance pay programs (300 schools) had test scores that were 0.35 standard deviations higher in language, and 0.54 standard deviations higher in math, than the scores of students in the control group (100 schools). *These effects are roughly equivalent to one additional year of schooling, which is very significant.*

In sciences and social studies, which were not covered by the program, test scores increased by 0.52 and 0.3 standard deviations higher than for control group students. These gains outside of

the program suggest that the effect of performance pay on teacher effort went beyond teaching to the test. *Individual teacher incentives were found to be more effective than group incentives over the long term.* Participating teachers would receive 500 rupees (US\$9.20) for every percentage point of average gain in the test scores of their students.⁵

Muralidharan concluded that teacher bonuses tied to learning outcomes were 15 to 20 times more effective than spending money on other improvements. Two clear lessons can be learned from this study: (1) The rules governing teacher incentives should be transparent, specifying clearly the rewards attached to increased efforts; and (2) the complementarities that exist among teaching to the test, designing broader tests, and instituting good teaching practices should be exploited.

4.4 School report cards

Andrabi, Das, and Khwaja (2009) assessed the market-wide impacts of providing report cards with learning scores to parents in 112 randomly chosen villages in Punjab province, Pakistan. One half of the villages were selected at random to be in the control group. The experiment covered 823 schools and 12,000 children in third grade, a total of 5,000 teachers, and background information on a sample of 1,800 households. *The provision of report cards increased test scores by 0.10 standard deviations and reduced private school fees by 23%.* However, there was wide variation across schools. In general, good private schools did not increase test scores, while government schools saw an increase of 0.10 standard deviations. Fee reductions were observed mostly in the good private schools.

An interesting caveat to the findings of this experiment is the reaction of private schools to the report card. The authors made two important points: First, the variation in student performance among public school students was much higher than the variation among private school students. Public schools showed proportionally higher gains from report cards because it was easier to get higher proportional gains among low-performing students in public schools than among better students in private schools. Second, among private schools, the tendency was for bad private schools to improve quality to keep market share, and for good private schools to reduce fees while maintaining quality. In the authors' judgment, the strategy of good private schools was reasonable because the additional cost of increasing quality among good private schools was higher than the potential loss of revenues resulting from lowering fees.

The report card included information on the academic performance of children and their relative performance with children in other schools. The report card was designed in collaboration with parents and schools. Parents wanted to know (1) their child's score and his/her rank relative to other students in the class, (2) the average score for each school in their village, and (3) the scores by category (word recognition, sentence building, etc.) so they would know where their child needed more help.

⁵ For details on the design of the program and the estimation of the teacher bonus, consult pages 7–8 of Muralidharan 2012.

Because many parents were illiterate, the report cards were discussed in small groups and the content of each card explained to the respective parents. The group also discussed what could be done to help, rather than to emphasize the negative or blame the child. The facilitators gave no advice during the discussions, to allow parents to come up with their own ideas.

Another case study on the use and impact of report cards on early grade reading comes from Liberia, where a randomized controlled trial (under EdData II) was conducted in 180 schools (Piper and Korda 2011). A group of 60 schools received a *full* treatment that included the Early Grade Reading Assessment among randomly selected students in the second and third grades, along with teacher training in the assessment of reading performance, frequent pedagogical support, and books and pedagogical materials. In addition, full-treatment schools produced for each student a color-coded (green, yellow, red) reading report card that was given to parents and a school report card given to the community. A second group of 60 schools received a *light* treatment, in which only the school report cards based on the Early Grade Reading Assessment were distributed among parents and the community. The control group of 60 schools did not get any of the above interventions.

The results of the experiment yielded a substantial increase in reading ability among full-treatment students and a substantive impact on the rate of progress in reading ability. Students in the *light* treatment group showed modest improvements over the control group in two of the seven items being measured. That is, the reading report cards for schools and students had some impact on reading performance.

The results suggest that informing parents about school quality and effectiveness on a regular basis makes a positive difference in student performance, although such impact may be modest. It must be noted, however, that although report cards may have a positive impact on learning, their innate purpose is to enhance accountability, not student learning; they should be expected to have an indirect effect on learning outcomes.

4.5 Parent participation in school management

The results of a randomized experiment on community-monitoring interventions in primary schools in Uganda showed that *students in schools with strong oversight by school management committees performed better than students in the control group* (Barr et al. 2012). Strong oversight of schools by school management committees seemed to improve test scores and reduce teacher absenteeism, and the differences between schools in the control group were statistically significant.

All schools had scorecards. However, in the control group, school management committee members received training and support in the use of scorecards developed by the Ministry of Education in collaboration with nongovernmental organizations. In the treatment group, school management committee members designed their own scorecard, defined their own objectives, and established their own indicators of progress. The treatment was called Participatory Scorecards. The authors concluded that participatory scorecards helped engage teachers and parents more than the Ministry's scorecard.

4.6 Accountability and education management information systems (EMIS)

It should be expected by now that accountability depends heavily on the availability and quality of information on school and system performance. One of the best systems in operation is in Chile (Ramírez 2012). Chile's *Sistema de Medición de la Calidad de la Educación* (SIMCE) is the main institution in charge of assessing and reporting on student learning outcomes. Created 20 years ago, the system is on par with international standards, with the purpose of informing policy makers, assisting in policy design, providing pedagogical support to teachers, and holding schools accountable. Each of these purposes is linked to different mechanisms for structuring and managing the dissemination of information.

SIMCE was not created and developed in a vacuum, as it benefitted from a favorable political moment that led to political support, a good institutional design, political pressure to perform, and the use of international experience as a frame of reference. However, the most important factor was the creation and approach of the SIMCE Commission, a dedicated group that generated the consensus needed to give the assessment system political and technical legitimacy.

SIMCE measures student achievement in four areas—math, language, natural sciences, and social sciences—assessing half a million students in 9,000 schools every year, with the assistance of 22,500 test supervisors and external administrators. Yearly results are front-page news and are discussed widely. The results are used to give feedback on policy design and evaluation, to devise teacher and school incentives, and—to a lesser extent—to make pedagogical changes (see Table 3, abridged from Ramírez 2012).

Table 3. Chile's dissemination strategy for its education assessment system

Assessment guidelines (since 1988)
Purpose: Pedagogical support. Audience: School principal, pedagogical coordinators, and teachers. Content: (a) Assessment framework and its relationship to the national curriculum. (b) Examples of test questions with an analysis of the contents and the skills required to answer them correctly. Distributed to all schools before the assessment (usually in the middle of the school year) and available online. Publication highly valued by teachers.
School report (since 1988)
Purpose: Pedagogical support. Audience: School principal, pedagogical coordinators, and teachers. Content: (a) National-, school-, and class-level mean scores by subject areas and grades tested. (b) Differences between school mean scores and mean scores from the previous assessment, from the national mean, and from schools of the same socioeconomic group. (c) Percentage of students by performance level—advanced, intermediate, beginner. (d) Examples of test questions with an analysis of the contents and skills required to answer them correctly. (e) Workshop guidelines for the schools to analyze assessment results and set improvement plan. Distributed to all participating schools at the beginning of the next school year.
National report (since 2006)
Purpose: Inform policy. Audience: Decision makers, general public. Content: (a) National and regional mean scores in subject areas and grades tested. (b) Percentage of students by performance level—advanced, intermediate, beginner. (c) Mean scores by socioeconomic background, gender, and by public/private school. (d) Trends in mean scores across years. Distributed at the central, regional, and provincial offices of the Ministry of Education, and to opinion makers.

<p><i>Newspaper supplement (since 1995)</i></p> <p>Purpose: Hold schools accountable. Audience: Parents, general public. Content: (a) School mean scores and mean scores by subject areas and grades tested. (b) Differences between school mean scores and mean scores from the previous assessment, from the national mean, and from the mean of schools from the same socioeconomic group. Published in newspapers with national and regional coverage; it usually includes school rankings.</p>
<p><i>Parent report (since 2005)</i></p> <p>Purpose: Hold schools accountable and involve parents in school. Audience: Parents. Content: (a) School mean scores, and mean scores by subject areas and grades tested. (b) Differences between school mean scores, and between subject area/grade mean scores of schools from the same socioeconomic group. (c) Percentage of students reaching different performance standards. (d) Recommendations to support student learning. Distributed to parents at the beginning of the school year. Also available online.</p>
<p><i>Press kit (since 2006)</i></p> <p>Purpose: Inform policy. Audience: Journalists and regional offices of education. Content: PowerPoint presentation with main results. Distributed to journalists during the press release.</p>
<p><i>Data files (since 2005)</i></p> <p>Purpose: Inform policy, provide pedagogical support, and hold schools accountable. Audience: Researchers. Content: Data files with student and school-level results. Provided upon request with a research proposal and a pledge to protect identity of students and teachers.</p>
<p><i>Data analysis tool (since 2007)</i></p> <p>Purpose: Inform policy, provide pedagogical support, and hold schools accountable. Audience: Researchers and decision makers. Content: Computes mean scores, differences in mean scores, and percentage of students reaching different performance standards. Computes results for different years, grades, public and private schools, subject areas, gender, and socioeconomic level, among others. Files with student-level data available upon request. Research proposal must have a pledge to protect identity of students and teachers.</p>
<p><i>Geo-referenced system (since 2010)</i></p> <p>Purpose: Hold schools accountable. Audience: Parents, so they can compare the scores of schools that are in the same geographic area. Content: Google maps with the geographical location of schools and their mean scores.</p>
<p><i>Website www.simce.cl (since 2001)</i></p> <p>Purpose: Inform policy, provide pedagogical support, and hold schools accountable. Audience: General public. Content: All the mechanisms described above, plus technical documents, and references to publications using SIMCE data.</p>

Source: Abridged from Ramírez 2012, p. 7.

Although countries can define their own strategies for disseminating their assessment results, Chile's system can be used as a design reference for countries without a good system for reporting assessment results. The question is: Does it work? A key problem in evaluating the

impact of information on learning outcomes is that the very nature of information on school performance generates a student selection bias, where the better schools attract the best students. This is a problem in Chile because the country relies on a voucher system for enrollment in privately managed schools. The econometric evidence so far indicates that, aside from self-selection biases, test scores tend to fluctuate widely from one year to the next, even when schools have maintained their input mix constant. Moreover, it is suggested that self-selection and the volatility of the school rankings are linked, which in turn indicates that producing school rankings that reflect reality is difficult (Mizala, Romaguera, and Urquiola 2007).

SIMCE is used as the data source for assigning monetary awards for teachers and schools that perform well on their tests relative to other schools in their comparable group. Analyses based on SIMCE information have suggested that the award program has increased language scores between 9 and 23 percent of a standard deviation, and math scores between 14 and 25 percent of a standard deviation (Rau and Contreras 2011). A previous analysis had failed to find a significant impact, corroborating the caution made earlier on the lagged impact of educational programs (Mizala and Urquiola 2007; Slavin 2013).

4.7 Education financing and system-wide incentives

Recently the World Bank has been using system-wide incentives to promote better performance in its educational loans. The idea behind these incentives is that countries know how to use positive and negative pressure within their educational system to ensure that education outcomes meet agreed-upon targets. If targets are not met, the additional funding is not forthcoming. The evidence with loans with Disbursement-Linked Indicators or with Performance-Based Contracts at the World Bank seems to be positive, but it is too early to know whether the approach has had an attributable impact on education quality and learning outcomes. However, the approach has an intuitive appeal that could be transferred to the school level, or at the subnational level. Table 4 presents a hypothetical example of indicators that could be linked to funding or to grants, using tourism-style pricing indicators to show relative disbursement amounts.

Table 4. Example of finance-linked indicators

Indicator	Baseline value	Target value	Source of verification	Disbursement amount if meeting 100% of target
Percentage of students in grades 1–3 reaching a standard that allows for comprehension: at least 40 wpm	15	50	Direct testing of student sample	\$\$
Percentage of students in grades 1–3 answering at least 80% of reading comprehension questions correctly	30	60	Direct testing of student sample	\$
Percentage of children in grade 1 able to recognize all letter sounds	5	50	Direct testing of student sample	\$\$\$
Percentage of school principals trained in management and leadership	0	30	Administrative data	\$

Indicator	Baseline value	Target value	Source of verification	Disbursement amount if meeting 100% of target
Percentage of teachers using scripted lessons for grades 1-3	0	80	Third-party verification of teacher sample	\$\$
Grade 6 student test scores in reading	350	380	Assessment system data	\$\$\$
Grade 6 student test scores in math	420	450	Assessment system data	\$\$\$

As in the case of test-based accountability, system-wide incentives can create distortions in the normal teaching practices of a school, inducing teachers to pay more attention to the indicators linked to additional funding. And, as in the case of test-based accountability, the choice of indicators, the reliability of the baseline score, and the feasibility of the target, are all variables that can be worked on to reduce distortions and to foster education quality.

Based on the experience to date, as noted by World Bank staff in personal interviews, the use of finance-linked indicators is a sort of back door to accountability, since getting additional funds requires that the borrowing country recognize its baseline indicators for learning, develop clear learning targets, and deliver results; otherwise there is no additional financing. Such an approach is deemed significant because it fosters a culture of incentive management and accountability at the highest levels within the educational bureaucracy (perhaps under pressure from their Ministry of Finance), which in turn forces education policy makers to develop the structure of support demanded by schools so they can be accountable.

4.8 Accountability and corruption in education

There is plenty of evidence showing that learning outcomes also depend on the proper functioning of educational institutions, and sometimes those institutions do not work as well as they should, because of corruption (Hallak and Poisson 2006). In education, corruption issues include low integrity, unethical behavior, and lack of transparency in several areas of system management:

- Education financing
- Teacher appointments and transfers
- Personnel conduct
- School nutrition programs
- Private tutoring

These operational areas of education can have corruption problems because of failures in the regulatory system or its enforcement, in the managerial capacity for compliance and enforcement, and in the lack of school-performance oversight by clients—especially parents.

In the volume edited by Hallak and Poisson, several country examples show the magnitude of the problem and its relation to systems of accountability. In most cases, the contributors to this volume described what was in place, but offered little numerical evidence to support assertions. This should be expected, as quantifying corruption is extremely difficult. However, going into more detail, the volume provides examples of countries—including Cambodia, Jordan, Lesotho, Mongolia, and Uganda—where corruption has been a problem in standardized testing; tutoring; student grading; and payoffs for graduation, school admissions, and teaching jobs.

In the cases of Argentina and Honduras, Morduchowicz (2006) and Arcia (2013) described education systems where the number of teaching positions exceeded the number of teachers required by school enrollment guidelines, with “ghost teachers” accounting for the oversupply. Some positions were being intentionally left unfilled, allowing teachers to have to multiple posts and draw more than one salary. The authors noted corruption in the management of personnel leaves, which were abused frequently, reducing class time and increasing effective costs; and “double dipping” in time sheets that were not verified. Argentina also was reportedly suffering from the creation of myriad job descriptions for specialized tasks, allowing for the hiring of many specialists to do little actual work.

In Colombia, corruption in teacher hiring was reduced substantially by the implementation of an education management information system that allows data on teacher appointments to be verified by parents (Peña and Rodriguez 2006). In some countries, such as Bangladesh, Nepal, and India, just specifying and implementing a clear code of conduct helped deter unethical behavior among teachers and staff (Khandelwal 2006).

Corruption also exists in school finance, but it can be reduced with the aid of formula funding. Evidence from England, Brazil, Poland, and the State of Victoria, in Australia, indicates that formula funding does work as long as some simple rules are followed (Levačić 2006). The main conclusion from these case studies on financial corruption was that there should be consequences for bad behavior, but these are not always applied.

5 Incentives and Accountability: A Summary of Key Lessons Learned

5.1 Practical lessons learned for implementing incentives and accountability

From the literature reviewed, one can glean *five necessary conditions for implementing incentives and accountability in education*:

1. The existence of clear performance standards aligned with the expected outcomes
2. An effective school support infrastructure
3. Enough autonomy for selecting and combining those inputs deemed essential to student outcomes
4. Information for evaluating school compliance with the standards

5. The application of rewards and sanctions resulting from school and teacher performance.

These can be further expanded:

Clear performance standards and goals. Learning standards should be defined by grade level at a minimum, and goals should be highly specific. Provision standards should be based on formula funding, which should be transparent for allocating funds at the regional and school levels. Standards should also be applied to teacher credentials and teacher behavior. Assessment systems should be implemented and a culture of assessment developed to replace the current vicious cycle of fear of assessment and the lack of progress tied to the lack of assessments.

Development of a culture of accountability. Accountability should be developed using parental empowerment as the main vehicle for implementation. Parent councils should be trained and given oversight power over school budget management, over personnel management, and over the application of learning standards.

Increased support at the school level. Teachers need to be prepared to become accountable, but in-service training should go beyond the traditional methods and become child-centered, not teacher-centered. In-service training should focus on mentoring and on direct orientation that would assist teachers to get children to reach the learning goals for the year.

The underlying issue that can be gleaned from the literature is that *accountability is intimately tied to the assessment and measurement of learning outcomes, to the incentive structure faced offered to teachers, to the management of assessment information, and to the expectations about the consequences for bad behavior.*

Assessment of student learning. Accountability requires information. To foster change, information must be understandable to parents and paired with the power and ability to leverage it for accountability. In terms of action, this may mean that:

- School administrators must be trained to interpret data on teacher or student performance, and must have the authority to respond; alternatively, inspectors and coaches from outside of the school need to be trained to use the data on learning outcomes and provide assistance to the schools.
- Teachers must be given the freedom to use student assessment information to adjust their teaching methods and trained on how to use it for the benefit of student learning; or asked to use a standardized or scripted approach if they prove incapable of improving results.
- Parents must also be trained to interpret assessment information and must have the power to hold school administrators and teachers accountable.

Teacher incentives. Expectation of change must be tied to an incentive of some kind—whether a salary bonus, special public recognition, community esteem based on results, or contract renewal. Intrinsic motivation is ideal, but cannot be predicted or counted on. For incentives to be effective, teachers must be convinced that the reward is tied to outcomes that are within their direct control, rather than on outcomes out of their control. Also, teachers must be able to predict the outcome of their efforts on the path to a reward. If teachers sense that their efforts will not work, or will be measured improperly, or will otherwise not be taken into account, the incentive

structure will not work. Neither will it work if teachers sense that the reward will come no matter what they do.

In summary, incentives and accountability measures are most successful in improving learning outcomes when enacted as part of an overall system reform program. Accountability measures can take the form of participatory school-based management or, as the European model shows, can be based on trust in schools and institutions. Where trust and the “long route” do not work, parents have a critical role in making accountability reforms work for positive learning outcomes. As stakeholders, parents’ involvement is more effective when they have access to good-quality information that they can understand, and that they can use to demand accounts from service providers and policy makers. There is therefore a need to create mechanisms that enable and empower parents and teachers to use the information they have received. Finally, accountability and incentives measures must be aligned with policy and standards in order to be effective.

In terms of accountability *mechanisms*, the empirical evidence indicates the following:

- In low-performing education systems, the increased participation of democratically elected school management committees is beneficial for establishing a culture of accountability and for setting the stage for improved learning outcomes.
- To make parent participation effective in fostering accountability, school autonomy in budget and personnel management must be increased.
- Training and support of school management committees is imperative in the management of financial resources, and in the interpretation of student assessments.
- The use of results of student assessment must be a collaborative effort between schools and the school management committees, to create a clear set of mutual expectations, and agreed-upon rewards and sanctions.
- School report cards, the simplification of outcome information for the benefit of both teachers and parents, and the use of learning outcomes as inputs for school management and pedagogy are crucial for ensuring that accountability is aligned with all other factors.

5.2 Generalized lessons for incentives and accountability

Educational incentives seem to respond to classic cost-benefit analysis at the institutional and personal levels. Incentives work well at these two levels when the cost of capturing them is lower than the value of the incentives themselves. That is why it is easier for students to respond to input-related incentives (conduct, uniforms, attendance) than to incentives related to outcomes, such as improved test scores. Some educational incentives related to learning outcomes assume that the school, teachers, or students know what to do to improve test scores. This assumption may not hold in most poor countries. To make teachers and schools accountable in terms of learning outcomes, the education system must also provide them with an infrastructure of assistance that operates all the way down to the classroom level.

The response to an incentive by teachers or students seems to have two components: a positive financial component—the value of the reward—and a negative component brought in by

uncertainty, which is the value of the probability of not performing well enough to get the reward. If the incentive rules are clear to all, and if the rewards are attached to attainable goals, the negative value of uncertainty decreases and the incentive is likely to work. This is important, as it underscores the need to have a support infrastructure aligned with incentives so that teachers and students can appraise the probability of success. If the support system is in place—such as teacher mentoring in the classroom, and access to adequate textbooks—the cost-benefit analysis made by teachers and students is bound to be favorable to incentives.

The experience to date shows that some incentives work and some do not, and success is highly specific to the school environment. As the evidence shows, monetary incentives seem to work well with inputs, but not with outcomes, and the magnitude of the incentives may vary greatly, along with the results. These results mean that the following general issues must be taken into account:

Many stakeholders with power. In education, the number of stakeholders with power may include politicians, community leaders, the Ministry of Education, the Ministry of Finance, funding agencies, and parent associations. Each of these stakeholders has different information needs, different expectations, and different views on the application of rewards and sanctions. Incentives and accountability must rely on common ground that takes into account the interests of these multiple stakeholders. That means that *the establishment of incentives and of procedures for accountability must be clear and agreed upon by these stakeholders.*

The possibilities for collusion exist among some stakeholders to extract rents from the education system. At the macro level, the Ministries of Education and Finance may use false baseline indicators to qualify for special financing from international agencies; at the mid-level, inspectors and schools may collude to affect enrollment statistics to get larger fiscal transfers; and at the school level, parents and schools may collude to report false enrollment or attendance data to get more funding for the school. These are examples of collusion already found in many countries. This means that entities promoting better educational performance through incentives and accountability must *ensure that the baseline data and education management information system are of good quality.* In fact, ensuring the implementation of a good EMIS may be a good goal in itself.

The interaction between accountability and the different levels of trust in governance must be understood. In countries with dysfunctional legal systems, trust is low. In these cases the short route to accountability may be a second-best solution for the proper functioning of education. The main lesson learned from the Finnish education system (Eurydice 2007) is that they have been successful in establishing a climate of trust, and *trust is very efficient*, as it implies very small transaction costs. This means that incentives and accountability in countries with low levels of trust (as described above) should rely on school-level actions and parent participation more than on bureaucratic mechanisms to manage trust at the local level. *Parent participation and school autonomy work well* in these contexts.

Deficiencies in the school support infrastructure. Strong accountability requires the education system to support schools in order to legitimately ask for accounts. If the support infrastructure is

absent or deficient, the best one can hope for is weak accountability—where information flows up from the school but rewards and sanctions are not applied. This strategy is common because political leaders cannot deliver on the conditions necessary for education agents to deliver good education services. Incentives and accountability must also have clear guidelines for the *provision of classroom and student support*.

Misalignment between policy and politics. In theory, politicians approve increases in teacher salaries in exchange for improvements in education quality. However, salary increases alone cannot produce higher teacher quality. Hence, as long as political leaders cannot negotiate merit pay in collective contracts, weak accountability will be more likely to prevail (Hanushek 2010). Financial incentives for teachers must have a clear link with teacher performance. If this link can be ignored, the incentive will not work as intended. Given this political reality, financial incentives for teachers are likely to work better if administered at the school level. Again, parent participation and school autonomy are good mechanisms for aligning policy with politics because decisions will be made at the school level by parents and the school.

The dissipation of accountability through the political process. In some countries, the responsibility for education management rests on education boards whose members are appointed, not elected. Using elections as the expression of parent voice is not very efficient, since it takes quite a long time for accountability to work (Hooge et al. 2012, p. 5). Even in cases of elected school board members, as in the United States at the local government level, accountability dissipates, making it very difficult to use elections to send clear messages to decision makers; that is why school board members in the United States tend to get reelected, since there is no one-to-one correspondence between bad school management and individual board members. The delegation of power to parents, and their integration into the school governance structure, give school autonomy and similar modes of school management a better chance for making incentives work as intended.

References

- Allan, Bradley M., and Richard G. Fryer Jr. 2011. "The power and pitfalls of educational incentives." The Hamilton Project Discussion paper 2011-07. Washington DC: Brookings.
- Andrabi, Tahir, Jishnu Das, and Asim Ijaz Khwaja. 2009. *Report Cards: The Impact of Providing School and Child Test-Scores on Educational Markets*. Working draft. Washington, DC: Development Research Group, World Bank.
- Arcia, Gustavo. 2013. *Análisis del financiamiento de la educación en Honduras* [Analysis of education finance in Honduras]. Consulting report. Washington, DC: Inter-American Development Bank.
- Arcia, Gustavo, Raja Bentaouet-Kattan, Harry Patrinos, and Angie Rivera-Olvera. 2013. *From Good to Great: The Impact of Programa de Escuelas de Calidad on Learning Outcomes in Mexico*. Washington, DC: Human Development Network-Education, World Bank.
- Arcia, Gustavo, Harry Patrinos, and Angie Rivera-Olvera. 2012. *SABER [Systems Approach for Better Education Results]—School Autonomy and Accountability. Mexico—Programa de Escuelas de Calidad (PEC)*. Human Development Network-Education. Washington, DC: World Bank
- Arcia, Gustavo, and Angie Rivera-Olvera. 2012. *Best Practices in Programa Escuelas de Calidad (PEC) Schools: Examples from Six Schools in Central Mexico*. Washington, DC: Human Development Network, World Bank.
- Arcia, Gustavo, Kevin Macdonald, Harry Anthony Patrinos, and Emilio Porta. 2011. *School Autonomy and Accountability*. Washington, DC: Human Development Network, Systems Approach for Better Education Results (SABER), World Bank.
- Arcia, Gustavo, Harry Patrinos, Emilio Porta, and Kevin Macdonald. 2011. *School Autonomy and Accountability in Context: Application of Benchmarking Indicators in Selected European Countries*. Washington, DC: Human Development Network, World Bank.
- Arcia, Gustavo, Emilio Porta, and Ramón Laguna. 2004. *Otro vistazo a la autonomía escolar de Nicaragua: Aceptación y percepción en 2004* [Another look at school autonomy in Nicaragua: Acceptance and perceptions in 2004]. Managua, Nicaragua: Consulting report, Programa de Promoción de la Reforma Educativa en América Latina y el Caribe (PREAL), CARE, UNICEF, and Ministry of Education.
- Arcia, Gustavo, and Humberto Belli. 1999. *Rebuilding the Social Contract: School Autonomy in Nicaragua*. LCSHD [Department of Human Development, Latin America and Caribbean Regional Office] Paper Series #40, Dept. of Human Development, Latin America and the Caribbean Regional Office, World Bank.
- Barr, Abigail, Frederick Mugisha, Pieter Serneels, and Andrew Zeitlin. 2012. *Information and Collective Action in the Community Monitoring of Schools: Field and Lab Experimental*

- Evidence from Uganda*. Working paper. Oxford, UK: Department of Economics, University of Oxford.
- Blimpo, Moussa, and David K. Evans. 2011. *School-Based Management and Educational Outcomes: Lessons from a Randomized Field Experiment*. Working paper. Washington, DC: Office of the Chief Economist, Africa Region, World Bank.
- Bold, Tessa, Mwangi Kimenyi, Germano Mwabu, Alice Ng'ang'a, and Justin Sandefur. 2013. *Scaling-Up What Works: Experimental Evidence on External Validity in Kenyan Education*. Washington, DC: Center for Global Development.
- Bovens, Mark, Thomas Schillemans, and Paul 'T Hart. 2008. "Does Public Accountability Work? An Assessment Tool." *Public Administration*, 86 (1): 225–242.
- Braun, Henry. 2005. *Using Student Progress to Evaluate Teachers: A Primer on Value-Added Models*. Policy Information Perspective. Princeton, NJ: Educational Testing Service.
- Bruns, Barbara, Deon Filmer, and Harry Anthony Patrinos. 2011. *Making Schools Work: New Evidence on Accountability Reforms*. Washington, DC: World Bank.
- Clarke, Marguerite. 2012. *What Matters Most for Student Assessment Systems: A Framework Paper*. Working Paper No. 1, Systems Approach for Better Education Results (SABER). Washington, DC: Human Development Network-Education, World Bank.
- Crouch, Luis, and Donald Winkler. 2007. "Governance, Management, and Financing of Education for All: Basic Frameworks and Case Studies." Paper commissioned for the *EFA [Education for All] Global Monitoring Report 2009, Governance, Management and Financing of Education for All*. Research Triangle Park, NC: RTI International.
- Di Carlo, Matthew. 2012. "Assessing Ourselves to Death." Blog post in Shanker Blog, <http://shankerblog.org/?p=6835>.
- Di Gropello, Emanuela, 2004. *Education Decentralization and Accountability Relationships in Latin America*. World Bank Policy Research Working Paper 3453. Washington, DC: World Bank.
- Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2012. "School Governance, Teacher Incentives, and Pupil-Teacher Ratios: Experimental Evidence from Kenyan Primary Schools." National Bureau of Economic Research (NBER) Working Paper No. 17939. Cambridge, MA: National Bureau of Economic Research.
- Elmore, Richard. 2003. *Knowing the Right Thing to Do: School Improvement and Performance-Based Accountability*. Washington DC: National Governors Association Center for Best Practices.
- Ewing, John. 2011. "Mathematical Intimidation: Driven by Data." *Notices of the American Mathematical Society* 58 (5): 667–673.
- Eurydice. 2007. *School Autonomy in Europe: Policies and Measures*. Brussels: Eurydice.

- Florida Department of Education. 2013. Grading Florida's Public Schools 2013. Retrieved from <http://schoolgrades.fldoe.org/pdf/1213/Guidesheet2013ShoolGrades.pdf>
- Figlio, David N., and Maurice E. Lucas. 2000. *What's in a Grade? School Report Cards and House Prices*. National Bureau of Economic Research (NBER) Working Paper No. 8019. Cambridge, MA: National Bureau of Economic Research.
- Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M. J. Vermeersch. 2011. *Impact Evaluation in Practice*. Washington DC: The World Bank.
- Glazerman, Steven, Susanna Loeb, Dan Goldhaber, Douglas Staiger, Stephen Raudenbush, and Grover Whitehurst. 2010. *Evaluating Teachers: The Important Role of Value-Added*. Washington, DC: Brown Center on Education Policy, Brookings Institution.
- Glewwe, Paul, Nauman Ilias, and Michael Kremer. 2010. "Teacher Incentives." *American Economic Journal: Applied Economics* 2 (3): 205–227. Retrieved from <http://www.aeaweb.org/articles.php?doi=10.1257/app.2.3.205>
- Goetz, Anne Marie, and Rob Jenkins. 2002. *Voice, Accountability and Human Development: The Emergence of a New Agenda*. Background Paper for the Human Development Report 2002. New York: United Nations Development Program.
- Gove, Amber, and Peter Cvelich. 2011. *Early Reading: Igniting Education for All. A Report by the Early Grade Learning Community of Practice* (rev. ed.). Research Triangle Park, NC: Research Triangle Institute. Retrieved from <http://www.rti.org/pubs/early-reading-report-revised.pdf>
- Hallak, Jacques, and Muriel Poisson, eds. 2006. *Governance in Education: Transparency and Accountability*. Paris: International Institute for Educational Planning.
- Hamilton, Laura S., Brian M. Stecher, and Stephen P. Klein, eds. 2002. *Making Sense of Test-Based Accountability in Education*. Santa Monica: RAND Corporation.
- Hanushek, Eric. 2010. *The Economic Value of Higher Teacher Quality*. Working Paper No. 56. Washington, DC: National Center for the Analysis of Longitudinal Data in Education Research, Urban Institute.
- Hanushek, Eric A., Suzanne Link, and Ludger Woessmann. 2013. "Does School Autonomy Makes Sense Everywhere? Panel Estimates from PISA." *Journal of Development Economics* (104): 212–232.
- Hermans, Peter, and Jan Wiegers. 2011. *The Use and Misuse of Assessment Data in Test-Based Accountability Systems*. Presented at the 37th International Association for Educational Assessment Annual Conference, Manila, Philippines.
- Hill, Carolyn J., Howard S. Bloom, Allison R. Black, and Mark W. Lipsey. 2007. *Empirical Benchmarks for Interpreting Effect Sizes in Research*. MDRC Working Papers on Research Methodology, New York: MDRC. Retrieved from <http://www.mdrc.org/publication/empirical-benchmarks-interpreting-effect-sizes-research>

- Hooge, Edith, Tracey Burns, and Harald Wilkoszewski. 2012. *Looking Beyond the Numbers: Stakeholders and Multiple School Accountability*. OECD Education Working Paper No. 85. Paris: OECD Publishing. Retrieved from <http://dx.doi.org/10.1787/5k91dl7ct6q6-en>
- Hout, Michael, and Stuart W. Elliott, eds. 2011. *Incentives and Test-Based Accountability in Education*. Washington, DC: The National Academies Press.
- Jerald, Craig D. 2012. *On Her Majesty's School Inspection Service*. Washington DC: Education Sector.
- Kane, Thomas J., and Douglas O. Staiger. 2002. "The Promise and Pitfalls of Using Imprecise School Accountability Measures." *Journal of Economic Perspectives*, 16(4): 91–114.
- Khandelwal, B. P. 2006. "Teachers' Codes of Practice in South Asia: Comparative Analysis of their Design, Implementation and Impacts in Bangladesh, India (Uttar Pradesh) and Nepal." In *Governance in Education: Transparency and Accountability*, edited by Jacques Hallak and Muriel Poisson, 154–169. Paris: International Institute for Educational Planning.
- Koretz, Daniel. 2008. *Measuring Up: What Educational Testing Really Tells Us*. Cambridge, MA: Harvard University Press.
- Kremer, Michael, Sylvie Moulin, and Robert Namunyu. 2003. *Decentralization: A Cautionary Tale*. Poverty Action Lab Paper No. 10. Cambridge, MA: Massachusetts Institute of Technology.
- Levačić, Rosalind. 2006. "Formula Funding of Schools, Decentralization, and Corruption: A Comparative Analysis." In *Governance in Education: Transparency and Accountability*, edited by Jacques Hallak and Muriel Poisson, 186–207. Paris: International Institute for Educational Planning.
- McDonnell, Lorraine M. 2002. "Accountability as Seen Through a Political Lens." Chapter 4 in *Making Sense of Test-Based Accountability in Education*, edited by Laura S. Hamilton, Brian M. Stecher, and Stephen P. Klein, 79–100. Santa Monica: RAND Corporation.
- Meza, Darlyn, J. L. Guzmán, and L. de Varela. 2006. *EDUCO: Escuelas Administradas por la Comunidad en las Zonas Rurales de El Salvador (1991–2005). Experiencias Latinoamericanas para Promover la Educación para Todos* [EDUCO: Community-managed schools in rural zones of El Salvador (1991–2005). Latin American experiences in promoting education for all]. Cartagena de Indias: World Bank/Inter-American Development Bank.
- Ministerio de Educación, Cultura y Deportes [Spain]. 2013. *Proyecto de Ley Orgánica para la Mejora de la Calidad Educativa* [Plan for the Organic Law to Improve Education Quality]. Madrid. This law was passed on May 17, 2013.
- Mizala, Alejandra, Pilar Romaguera, and Miguel Urquiola. 2007. "Socioeconomic Status or Noise? Tradeoffs in the Generation of School Quality Information." *Journal of Development Economics* 84: 61–75.

- Mizala, Alejandra, and Miguel Urquiola. 2007. *School Markets: The Impact of Information Approximating Schools' Effectiveness*. National Bureau of Economic Research (NBER) Working Paper 13676. Cambridge MA: National Bureau of Economic Research.
- Morduchowicz, Alejandro. 2006. "Efficiency and Corruption in Education: The Case of Argentina." In *Governance in Education: Transparency and Accountability*, edited by Jacques Hallak and Muriel Poisson, 82–101. Paris: International Institute for Educational Planning.
- Mourshed, Mona, Chinezi Chijioke, and Michael Barber. 2010. *How the World's Most Improved School Systems Keep Getting Better*. London: McKinsey & Co. Retrieved from <http://mckinseysociety.com/how-the-worlds-most-improved-school-systems-keep-getting-better/>
- Muralidharan, Karthik. 2012. *Long-Term Effects of Teacher Performance Pay: Experimental Evidence from India*. Working paper. San Diego, CA: Department of Economics, University of California–San Diego.
- Organization for Economic Co-operation and Development (OECD). 2013. *PISA 2012 Results: What Makes Schools Successful? Resources, Policies and Practices (Volume IV)*. Paris: OECD.
- OECD. 2011. "School Autonomy and Accountability: Are they Related to Student Performance?" *PISA in Focus* No. 9. Paris: OECD.
- Pearson Education. 2004. *Value-Added Assessment Systems*. Policy Report. San Antonio: Pearson Education.
- Peña, Margarita, and Jeanette Sofia Rodriguez. 2006. "Transparency in the Education Sector: Human Resource Management in Bogotá, Colombia (1998–2003)." In *Governance in Education: Transparency and Accountability*, edited by Jacques Hallak and Muriel Poisson, 102–141. Paris: International Institute for Educational Planning.
- Piper, Benjamin, and Medina Korda. 2011. *EGRA [Early Grade Reading Assessment] Plus: Liberia—Program Evaluation Report*. Prepared for USAID under the EdData II Project, Task 6, EHC-E-06-04-00004-00. Research Triangle Park NC: RTI International. Retrieved from <https://www.eddataglobal.org/documents/index.cfm?fuseaction=pubDetail&ID=283>
- Ramírez, María-José. 2012. *Disseminating and Using Student Assessment Information in Chile. Systems Approach for Better Education Results (SABER)*, Working Paper No. 3. Washington, DC: Human Development Network-Education, World Bank.
- Rau, Tomás, and Dante Contreras. 2011. *Tournament Incentives for Teachers: The Case of Chile*. Working paper. Santiago: Department of Economics, University of Chile.
- Reubens, Andrea. 2009. *Early Grade Mathematics Assessment (EGMA): A Conceptual Framework Based on Mathematics Skills Development in Children*. Prepared for USAID under the EdData II Project, Task 2, EHC-E-02-04-00004-00. Research Triangle Park,

- NC: RTI International. Retrieved from <https://www.eddataglobal.org/documents/index.cfm?fuseaction=pubDetail&id=193>
- Rosenkvist, Morten Anstorp. 2010. *Using Test Results for Accountability and Improvement: A Literature Review*. OECD Education Working Paper No. 54. Paris: Organisation for Economic Co-operation and Development.
- Sanders, William L., S. Paul Wright, June C. Rivers, and Gil G. Leandro. 2009. *Addressing Common Concerns About Value-Added Modeling*. A SAS White Paper. Cary, NC: SAS Institute.
- Slavin, Robert. 2013. "Do Clinical Trials Work in Education?" *The Huffington Post*, August 4. http://www.huffingtonpost.com/robert-e-slavin/do-clinical-trials-work-i_b_3681611.html
- Solomon, Pantelis. 2012. *Do Schools Learn: Response to North Carolina's School Accountability Program*. Job Market Paper. Providence, RI: Department of Economics, Brown University.
- Vegas, Emiliana, and Ilana Umansky. 2005. *Improving Teaching and Learning Through Effective Incentives. What Can We Learn from Education Reforms in Latin America?* Washington, DC: World Bank.
- West, Martin R., and Paul E. Peterson. 2006. "The Efficacy of Choice Threats Within a School Accountability System: Results from Legislatively Induced Experiments." *The Economic Journal* 116: C46–C62.
- World Bank. 2011. *Enhancing School-Based Management in Indonesia*. Human Development Sector Policy Brief No. 60457. Jakarta: World Bank.
- World Bank. 2003. *World Development Report 2004. Making Services Work for Poor People*. Washington, DC: World Bank.

Annex: Grading Formula for the Florida Comprehensive Assessment Test (FCAT)

School Classification under the Florida Comprehensive Achievement Test, Elementary and Middle Schools

Schools are awarded one point for each percent of full-year-enrolled students who score proficient or higher on the FCAT and who make annual learning gains (800 possible points).

Grading Scale: A = at least 525 points, B = 495 to 524 points, C = 435 to 494 points, D = 395 to 434 points, F = fewer than 395 points.

Performance Components: 50% of total. 400 points maximum.

Percent scoring Satisfactory or higher on FCAT Reading (max 100 pts.)

Percent scoring Satisfactory or higher on FCAT Mathematics(max 100 pts.)

Percent scoring Satisfactory or higher on FCAT Science (max 100 pts.)

Percent scoring Satisfactory or higher on FCAT Writing (max 100 pts.)

Learning Gains Components: 50% of total. 400 points maximum. Students achieve learning gains if (i) if they had a score of satisfactory or higher and increase their score by one or more achievement levels or (ii) were at levels 1 or 2 but demonstrate more than a year's growth on the FCAT developmental score scale.

Percent who made learning gains in reading (max 100 pts.)

Percent who made learning gains in mathematics (max 100 pts.)

Percent in the lowest performing 25% who made learning gains in reading (max 100 pts.)

Percent in the lowest performing 25% who made learning gains in math (max 100 pts.)

Schools must test at least 90% of eligible students (at least 95% to be eligible for an "A") to be assigned a regular letter grade; if not, they are graded "I" and are subject to penalties. School must show adequate progress of the lowest 25% in reading and math. For schools earning a grade of "C" or higher based on points, at least 50% of the lowest performing 25% in each subject must make learning gains or the school will be assigned a final grade one grade lower than it would otherwise have earned based on total points. Schools falling below the 50% criterion can still meet the requirement by showing annual improvement in the percentage of low-performing students making learning gains.

Source: Florida Department of Education. <http://schoolgrades.fldoe.org/pdf/1213/Guidesheet2013ShoolGrades.pdf>