CICE Current Issues in Comparative Education

Volume 14(1) / Fall 2011 • ISSN 1523-1615 • http://www.tc.edu/cice

Evaluation, Assessment, and Testing

FEATURED ARTICLE

BOOK REVIEW

CURRENT ISSUES IN COMPARATIVE EDUCATION
Volume 14, Issue 1 (Fall 2011)

Issue Editors:
Ruaridh J. MacLeod, Sarah Flatto

How Much is Learning Measurement Worth? Assessment Costs in Low-Income Countries

Daniel A. Wagner[1]
International Literacy Institute, University of Pennsylvania

Andrew Babson
International Literacy Institute, University of Pennsylvania

Katie M. Murphy
International Literacy Institute, University of Pennsylvania

Timely and credible data on student learning has become a global issue in the ongoing effort to improve educational outcomes. With the potential to serve as a powerful diagnostic tool to gauge the overall health and well-being of an educational system, educational assessments have received increasing attention among specialists and the media. Though the stakes are high, relatively little is known about the cost-benefit ratio of various assessments compared to other educational expenditures. This paper presents an overview of four major types of assessments—national, regional, international and hybrid—and the costs that each has incurred within 13 distinct contexts, especially in low-income countries. The findings highlight broad variation in the total cost of assessment and the cost-per-learner. This underscores the importance of implementation strategies that appropriately consider scale, timeliness, and cost-efficiency as critical considerations for any assessment.

Throughout the global arena, government agencies, international organizations, donors and private sector partners increasingly emphasize the need for timely and credible data on student learning that may inform the design of effective mechanisms to improve educational outcomes. Considerable attention at these multiple levels, compounded by a heightened publicity in the media, has prompted a dramatic and global growth in the use of learning assessments (Kamens & McNeely, 2010).

At the same time, the reality of restricted educational budgets demands affordable and cost-effective options for assessments. Indeed, with the growth of Large Scale Educational Assessments (LSEAs) throughout the world, there has been a concomitant increase in attention to the fiscal burden of assessments in low-income countries (LICs). These costs have often been borne by external funders such as bilateral or donor agencies, resulting in the common perception that this burden of investments in knowledge is rather minimal when compared to the large amounts spent on education itself (Lockheed & Hanushek, 1988; Porter & Gamoran, 2002).

The perception that LSEAs are relatively low-cost has been supported by a limited number of studies showing that assessment costs represent a very small proportion of national education budgets.[2] Yet these studies do not appear to account for the increasingly limited amount of discretionary funds for such activities that may be available to ministers of education in low-income countries, with or without external support. Hence, for more than a decade, other critical perspectives have emerged that challenge the assertion that LSEAs are a relatively small-scale investment.[3]

The actual costs of LSEAs and other assessments are needed in order to determine cost-benefit analyses within low-income countries. A successful international assessment requires high-level skills in design, planning and management—skills that are in short supply globally—especially in LICs.[4] Ministries of education throughout the world are now confronting difficult decisions in regard to assessments. First, they must decide whether to participate in LSEAs, understanding the costs and complexity of large-scale assessments. Second, they must determine how to choose tests that are appropriate for students, languages and educational systems[5] given the wide variety of assessments available today.

Ensuring that policymakers and key stakeholders have accurate information on the actual costs of assessments is a critical step in identifying appropriate tools to inform and influence initiatives aimed at improving educational outcomes. This paper presents an overview of four major types of assessments: national, regional, international and hybrid, and the costs that each of these types of assessments have incurred during implementation in various contexts. The findings highlight a broad variation in the total cost of assessment and the total cost-per-learner within each of the four types of assessments.

Types of Educational Assessments

Large-scale educational assessments (LSEAs) have increasingly been used by national and international agencies beginning in the 1980s. Previously, only a small number of cross-national large-scale assessments had been conducted, mostly by the IEA (International Association for the Evaluation of Educational Achievement).[6] Technological and methodological advances in assessment, combined with the political pressure to improve educational systems, have spurred this trend, including in LICs (Kelleghan & Greaney, 2001). The 1990 Jomtien Conference demanded more accountability and systemic evaluation of education in LICs. Further, in 2000, the UNESCO Dakar Framework for Action called for the achievement of "measurable" learning outcomes, and that such progress should be "monitored systematically" (UNESCO, 2000, p. 21). LSEAs have increasingly become a key tool for meeting these demands.

Despite this momentum, the increasing complexity and expense of LSEAs have led to questions about the utility of conducting LSEAs in low-income countries. Although a number of agencies have carried out LSEAs in the OECD (Organization for Economic Co-operation and Development) countries, it was not until the 1990s that the capacity to participate in LSEAs (international and regional) became more available to LICs. The complexity of stakeholder

interests, as well as resource constraints, has limited growth of LSEAs in LICs. However, various donor agencies, such as the World Bank, have become increasingly important funders of LSEAs, making it more affordable and more likely for such assessments to be utilized even when national budgets are constrained.[7]

With a focus on learning assessments in low-income countries, the present discussion centers on four main types of assessments: national, regional, international and hybrid. Each of these is described below.

National assessments

National assessments (sometimes called national or public examinations) evaluate all students in a national educational system. Nearly all countries engage in some type of national assessment in order to ascertain whether desired and planned educational goals are achieved. The results can be used to modify curricula, train teachers, reorganize school access, or refashion numerous other aspects of a national educational system. The results also can be used for accountability purposes, to make resource allocation decisions, and to heighten public awareness of education issues.

Regional assessments

Regional assessments provide an opportunity to measure student learning across a group of countries, typically defined by a geographic region or by a shared national language. They have grown in popularity over the last 20 years, and as part of an effort to extend the use of LSEAs into developing countries, regional and international organizations have collaborated to create three major regional assessments: the *Latin American Laboratory for Assessment of Quality in Education* (LLECE), the *Southern and Eastern African Consortium for the Monitoring of Education Quality* (SACMEQ), and *Program for the Analysis of Educational Systems of the CONFEMEN* (francophone Africa) countries (PASEC).

International assessments

International assessments are designed to measure learning in multiple countries. Their aims include: (a) cross-national comparisons that target a variety of educational policy issues; (b) provision of 'league tables' that rank-order achievement scores by nation or region or other variables; and (c) within-country analyses that are then compared to how other countries operate at a sub-national level. These studies are undertaken by various international organizations and agencies, including: the IEA that conducts the *Progress in International Reading Literacy Study* (PIRLS), and the OECD that is responsible for the *Program for International Student Achievement* (PISA) studies. Each of these international assessments is now in use in dozens of countries, and is expanding well beyond the OECD country user base that formed the early core group of participation.

Hybrid assessments

In recent years, a new approach to assessment has sought to focus more directly on the needs of LIC assessment contexts. Initially, this approach was conceptualized under the acronym

"smaller, quicker, cheaper" (SQC) methods of literacy assessment (ILI/UNESCO, 1998; Wagner, 2003). The idea was to see whether LSEA methodologies could be reshaped into *hybrid*[8] methods that are just big enough, faster at capturing, analyzing and disseminating data, and cheaper in terms of personnel and cost outlays (Wagner, 2010, 2011). The *Early Grade Reading Assessment*, or EGRA, (Research Triangle Institute, 2009) contains a number of the above features, and is probably the best-known current example of a hybrid assessment in reading. The EGRA was initially designed with three main assessment goals: early reading (grades 1-3), local contexts (rather than comparability across contexts), and local linguistic and orthographic variation.

Cost-benefit Analyses in Educational Assessment

In the early 1990s, a limited number of studies examined the costs and benefits of LSEAs (Ilon, 1992, 1996; Koeffler, 1991; Loxley, 1992; Lockheed, 2008). The results supported the value of LSEAs for two main reasons: the fairly low overt costs (i.e., costs that are explicitly budgeted and accounted for) in relation to the overall education budget (Peyser & Costrell, 2004; Hoxby, 2002), and the high potential benefits of LSEAs to yield actionable results (Braun & Kanjee, 2006; Hanushek & Woesmann, 2005). Nonetheless, as pointed out by Lockheed (2008, p. 9), "national learning assessments in developing or transition countries rarely employ complex measurement instruments because such countries rarely have the requisite domestic capacity or can afford to purchase expertise from abroad." This point is echoed by Greaney and Kelleghan, (2008) and Wolff (2007), and further linked to potential wastage or failure down the road if sufficient investments are not made up front (Wolff, 2008).

Thus, while the assessment field itself—whether in high- or low-income countries—seems largely convinced of the importance of LSEAs, the total costs of assessments are becoming more clearly recognized as a serious obstacle for LICs. For example, Braun and Kanjee (2006) assert that, in countries with developing educational systems, "scarce resources are better devoted to assessments directed at improving learning and teaching, where the returns on investments are likely to be higher."[9] Research shows that the average costs of an LSEA appear small relative to national educational budgets (less than 1% generally per national budget, and as low as 0.3%). However, such low percentages may not reflect the percentage of the available discretionary budget (Coombs & Hallak, 1987, p. 50).

Calculating the costs

In order to make a cost-based decision about assessment choice, it is important to bear in mind both overt and hidden costs that come into play in any assessment (cf. Greaney & Kellaghan, 2008). Overt costs are those that are typically planned for in advance and that are included in the accounting mechanisms of the agency (or agencies) in charge of the LSEA. These would include staff costs of test management (such as test design and application) and training, as well as travel, supplies and equipment.

They can also vary by location, including: within-country costs (e.g., roll out and management of the assessment process within country); in-kind costs (e.g., non-cash contributions such as ministry staff, specialists, headmasters, and teachers); and international costs (e.g., international agency overheads, international experts, and travel).

Hidden costs are those that may escape the attention of authorities that put together fiscal plans for assessments. They include the following items.

- *Indirect (or overhead) costs*. These costs are absorbed by the agencies themselves in implementing the program. While often accounted for in wealthier countries, these costs sometimes escape the attention of ministries and other agencies in LICs. Obvious examples would include the cost of using infrastructure (e.g., buildings, networks, computer maintenance, and so forth). Less obvious, but significant, costs may be associated with seconded staff in the ministry and field workers who may be school inspectors or teachers.

- *Opportunity costs*. These costs are relative to what different strategy may have taken place in lieu of the particular choice that is made. For example, by not doing an assessment in a particular year, the ministry might have more resources to do the assessment in a subsequent year. Or, choice of one type of assessment may preclude opting for an additional or different choice.[10] However, the cost of *not* participating in an assessment—that is, foregoing the potential benefits (in terms of staff development, potential results, etc.) of participation in an assessment—must also be considered as another type of opportunity cost.

Cost Categories and Comparisons: Selected Assessments

The cost categories in assessments from the previous discussion may be seen in summary form in Appendix A. For purposes of comparison, a number of well-known assessment agencies were contacted for current information on expenditures (some in estimated form). The studies covered are listed in Appendix B. Data collected from each of the selected studies at a national level are represented in Table 1, which indicates the variability of known assessment costs, by assessment and national context across 13 recent assessments. Table 2 provides a summary of average percentages of total expenditures across the six main cost categories.[11]

As shown in Table 1, it is possible to make a number of observations. First, the student populations ranged from a modest 3,770 in EGRA-Liberia, to about 300,000 in SIMCE (Chile).[12] Second, it may be seen that the total (listed) overt costs of undertaking the assessment range from a low of about $122,000 in PISA (Uruguay) to a high of $2.8 million in SIMCE (Chile). Third, by considering these first two parameters, it is possible to calculate the 'cost-per-learner' (CPL) assessed, a useful way of looking at costs irrespective of size of the total enterprise. Results indicate that cost-per-learner ranges from about $8 in the Uruguay national assessment to about $51 in the SACMEQ III study in Swaziland to about $171 in PISA in Chile. The average for this sample of studies is about $42 per learner assessed. In addition (see Table 2), certain costs figured more prominently than others, such as test application (50%) and

institutional costs (23%), while processing and analysis (13%) and test preparation (11%) were substantially lower.[13]

The average CPL data show that, at the field level, these are not dramatically different when compared across types of tests. Some assessments are clearly more expensive, but it is interesting to note that the larger national and international studies confer economies of scale that reduce per-unit assessment costs. At present, the smaller EGRA studies are not less expensive at the field level. Further, some countries may have significantly more resources (financial, intellectual, infrastructural, etc.) in their evaluation departments upon which to draw. This will likely affect a number of cost variables, such as in-house versus external consulting fees and travel expenses. It must be understood that hybrid assessments are still in a research phase (with inherent costs of trial and error), such that their costs may be expected to drop substantially with the establishment of economies of scale. In addition, specific in-country needs and requirements (e.g., logistics in difficult terrain) may also play a major role in determining which types of assessment are chosen, and thus how much is ultimately spent on assessment.

Of course, much depends on whether cost estimates are correct and whether hidden costs are fully included. Not all teams collect and store cost data and, even if they do so, these data may not be complete or sufficiently detailed for comparative analyses. Inaccuracies and discrepancies are often the result of underfunding (Lockheed, 2008, p. 16). Thus, these data should be considered a preliminary view of cost comparisons, and more needs to be done with full and reliable auditing in place.

Cost parameters with low-income countries in mind

In low-income countries, educational decision makers will find themselves with more choices than available resources. The cost-benefit picture remains insufficient. Simply not enough reliable data have been collected on assessment costs for the variety of assessments currently in use. Moreover, the current scientific, technological and political dynamism in educational improvement strongly suggests that models of assessment will change in relation to testing advancements and increasing demand. The necessity for both clear testing choices and actionable indicators is likely to increase.

Recent assessment innovations (e.g., EGRA) suggest momentum toward models of assessment that both emphasize a needs-centered and 'just enough' approach to testing (Wagner, 2003). This means that innovations may help to grow the scale of test application, shrink upfront overt costs such as translation and test preparation, and reduce turnaround time. This way, government bodies can possess actionable data sooner and thus with less staff and overhead. Three key parameters summarize the cost issues of assessments that will need to be considered, especially in the context of resource-constrained LICs.

Scale

Ministries of education in LICs will need to consider which assessments would yield targeted and responsive educational data about a specific population (e.g., rural girls, ethno-linguistic groups), a group of schools, or concerning a particular subject at a particular grade level. LSEAs typically cannot respond flexibly to such requests due to the significant up-front preparation and pre-assessment exercises that constrain near-term changes, and lock in comparability parameters. Further, most LSEAs are not designed to provide classroom-level indicators but rather systemic indicators (Volante, 2006). By contrast, limited sample household-based surveys or EGRA style hybrid assessments can save money because they can reduce the number of individuals to be assessed in order to answer a more specific set of policy questions, and can be deployed and adjusted more frequently. Still, recent sampling innovations in LSEAs (such as PIRLS) suggest that such studies not only provide multi-level data, but also that the economies of scale can enable larger samples at marginal additional cost.[14] In other words, lower cost (in CPL) is a relative term.

Timeliness

Two types of timeliness are crucial to the possible benefits of assessments. First, there is the timeliness of the testing cycle from planning, rollout, and data collection to analysis and dissemination (and subsequent policy debates). Second, timeliness can also refer to the 'right time' of information availability and use. For example, if timely information about a group of schools is ready in advance of major school finance decisions, then those data can show real-time sensitivity. Or, a population of students may need assistance to reach grade-level competence in reading, and data may confirm, disconfirm, and/or guide the decision-making process. In addition, there is a need to consider the merits of early intervention in the learning trajectory of students, much as the arguments have been made in the medical field for early detection systems.[15] In sum, credible assessment data needs to be gathered as quickly as possible in order to effectively shape policymaking, yet it also needs to be available for application to decision-making at the right time. If 'time is money' (as the adage goes), then moving toward timeliness can also help to reduce overall costs of assessment and intervention.

Cost efficiency

As mentioned above, some assessments are relatively expensive in terms of up-front cost outlays, with requirements of expensive professional staff and consultants, and trained field enumerators. These and other costs can be seen in terms of either total costs or the CPL. Either way, budgetary limits on discretionary funds in LICs will require careful scrutiny as assessment choices are made. Given the paucity of credible data on costs in LICs today, it is difficult to derive an evidence-based decision pathway for multiple contexts. There is a clear need to more precisely determine which expenditures are likely to reveal particular policy outcomes. For example, will increasing expenditures for the training of enumerators yield better inter-rater reliability? Or, as in a recent effort in India, can volunteers become low-cost, reliable and sustainable enumerators with relatively little training at all (Banerji, 2006)? More research is needed to better clarify the cost merits of different assessments.

Conclusions

Costs are an inherent part of any social intervention. The assessment of learning and its policy consequences constitute a clear case in point. The key issue here is not that assessments are "expensive" or not. Rather, the issue is what a ministry (or funding agency) will receive in return for its investments.

Gathering data on the *comparative* costs of assessments is dificult. There are, however, some reference points now available that can be considered. Perhaps most important is the trade-off between time and money. Take, for example, a minister of education who may have up to five years to decide upon and implement policy. In this case, regional or international LSEAs such as SACMEQ or PASEC may provide some solid answers on key issues, and offer a sense of cross-national comparison. Given the current economies of scale in countries that repeat international assessments, the actual CPL of such LSEAs is not much different from that of the EGRA and hybrid assessments that have much smaller sample sizes.

On the other hand, if a minister has a shorter window of policymaking opportunity (such as the typical two to three-year mandate in office), and if the priority is helping programs, schools and regional districts attain their near-term learning achievement goals, even a small-scale sample-based assessment like EGRA looks much less expensive. While the CPL in EGRA appears similar to the larger international assessments at present, the future costs will likely drop as EGRA tools become more familiar, enumerator training improves,and technological advancements reduce the amount of time and human resources required to analyze and disseminate assessment data.

Finally, there are opportunity costs to consider. LSEAs are typically not administered until children reach grade 4 (or later), when children may be far behind in reading development; this can impose very high costs in remediation that early assessment could prevent. "Catching up" is expensive, difficult, and may lead to school failure—the most important cost that policy makers seek to avoid.

In sum, evaluating and learning from assessments is fundamental to credible change in educational systems across nations. But learning assessments entail costs that need to be evaluated and compared. Gone are the days when ministerial agencies can assign free seconded staff to the field, or when outside donor agencies will fully fund large scale assessments. We are in a time of fiscal constraints. Learning about education has to be balanced against *what is learned*, *for what purposes, and at what cost*. The evaluation of assessment costs is an issue that will need considerably greater attention in the field of international education.

Table 1

Test monetary costs (USD)	National Assessments			Regional Assessments			International Assessments					EGRA Assessments[i]	
	SIMCE 2004[a]	Honduras 2004[b]	Uruguay 2003[c]	PASEC 2010[d]	SACMEQ III Swaziland 2007[e]	SACMEQ III Tanzania 2007[f]	PISA Chile 2009[g]	PISA Mexico 2009[h]	PISA Panama 2009[i]	PISA Peru 2009[j]	PISA Uruguay 2003[k]	EGRA - Liberia 2008	EGRA - Nicaragua 2008
Test preparation	**258,236**	**174,275**	**21,528**	**34,164**	**12,561**	**12,666**	**26,448**	**100,301**	**61,475**	**47,956**	**12,357**	**29,345**	**10,882**
Creation and editing of test items	184,515			7,895		1,000	26,448	3,802	13,661				
Pilot testing	73,721			15,749	12,561	11,666		96,499	47,814			16,031	4,756
Training				10,520								13,314	6,126
Test application	**1,163,764**	**435,717**	**57,289**	**91,705**	**170,732**	**89,900**	**597,958**	**891,501**	**187,157**	**212,486**	**29,707**	**82,260**	**68,683**
Test design and editing	29,403			7,415		2,000	8,976		13,661	2,590		8,800	
Test printing	324,712			9,744	15,488	12,000		254,899	54,644	7,196		5,600	1,395
Printing of other materials	236,076				3,049	4,200		116,156	6,831				
Distribution to examiners	103,124			6,455	73,171	2,000		123,845	6,831				
Field testing	406,103			68,091	79,024	56,700	462,705	394,235	98,359	198,261		67,860	67,288
Control and supervision	64,346					13,000	126,277	2,366	6,831	4,439			
Processing and analysis	**382,239**	**130,721**	**26,272**	**12,624**	**454**	**33,300**		**167,782**	**128,414**		**22,838**	**13,533**	**5,734**
Coding and digital input	216,048			12,624		33,300		56,899	114,753			13,533	5,734
Marking open-ended questions	166,191				454			110,883	13,661				
Additional analyses													
Dissemination	**100,567**	**130,721**	**531**	**32,193**	**4,195**	**2,000**	**49,912**		**34,153**	**3,865**	**14,092**	**1,850**	
School communication	100,567				4,195	2,000			34,153	3,865		1,500	
Report production and distribution							49,912					350	
Public relations retainer													
Subtotal	**1,904,806**	**871,434**	**105,620**	**170,686**	**187,942**	**137,866**	**674,318**	**1,159,584**	**411,199**	**264,307**	**78,994**	**126,988**	**85,299**
Institutional costs	**938,766**			**12,481**	**24,878**	**25,500**	**179,233**	**490,203**	**94,261**	**20,473**		**103,520**	**87,157**
Personnel- in project budget	796,864			2,737	17,561	10,000	179,233	321,246	73,769	9,324		101,858	83,675
Personnel- contributed								107,286		11,149		1,403	2,500
Infrastructure- in project budget	35,369					5,000		2,743	6,831				
Infrastructure- contributed													
Equipment in - project budget	106,533			9,744	7,317	10,500		58,928	13,661			259	982
Equipment- contributed													
Test Fees							**49,863**	**118,599**			**43,197**		
Other	**20,028**			**2,043**			**72,494**		**13,661**	**2,000**		**10,619**	**6,958**
TOTAL	**2,863,600**	**871,434**	**105,620**	**185,210**	**212,820**	**163,366**	**975,908**	**1,768,386**	**519,121**	**286,780**	**122,191**	**241,127**	**179,414**
Total Students	300,000	45,657	12,993	5,400	4,155	3000 [m]	5700 [n]	45,079	42,000	7,967	5,797	3,770	5,760
Total Schools												240	120
Cost per student	**10**	**19**	**8**	**34**	**51**	**55**	**171**	**39**	**12**	**36**	**21**	**64**	**31**
Cost of educating a student	767	130	484		66			9,439	1,023	396	479		
Cost of testing as % of total budget for one grade	0.83	2.63							1.20838				
Cost of testing as % of total secondary education budget	0.17	0.33	0.07					0.001767	0.04419		0.08		

Table One: Detailed costs for national, regional, international and EGRA assessments.

[a] Source: Wolff, 2007, p. 6 (for 2004 SIMCE test). Wolff (2007) used local currencies for his figures on PISA Uruguay 2003 and all the national assessments above (namely SIMCE 2004, Honduras 2004 and Uruguay 2003). In order to facilitate comparisons across assessments in this table, we converted Wolff's figures to the average annual market rate for USD. Further, in his analysis of SIMCE 2004, Wolff used SIMCE 2002 figures, in Chilean Pesos (At the rate of 677.4916667 Chilean Peso to 1 USD).

[b] Source: Wolff, 2007, p. 13; 2004 17.68 Honduran Lempira to 1 USD

[c] Source: Wolff, 2007, p. 11; 2003 28.24279 Uruguayan Peso to 1 USD

[d] Source: PASEC 2010 technical report (personal communication, P. Varly, May 2009). Converted from Euros to USD, 2009 annual rate.

[e] Source: Personal communication, A. Mrutu, August 2009.

[f] Source: Personal communication, J. Shabalala, August 2009.

[g] Source: Personal communication, E. Lagos, September and October 2009.

[h] Source: Personal communication, M. A. Diaz, September 2009.

[i] Source: Personal communication, Z. Castillo, September 2009.
[j] Source: Personal communication, L. Molina, September 2009.
[k] Source: Wolff, 2007, p. 14; 28.24279 Uruguayan Peso to 1 USD (2003)
[l] Source: Personal communication, A. Gove, August 2009.
[m] Estimate, based on SACMEQ II sample of 2854
[n] Estimate, based on email of E. Lagos, October 2009

Table 2.

Cost category	Average	Lowest	Highest
Test preparation	11%	3% (PISA Chile, 2009)	20% (Uruguay, national assessment, 2003)
Test application	50%	24% (PISA Uruguay, 2003)	80% (SACMEQ III, Swaziland)
Processing and analysis	13%	1% (SACMEQ III, Swaziland)	25% (Uruguay, national assessment, 2003)
Dissemination	6%	1% (Uruguay national assessment, 2003)	17% (PASEC, 2010)
Institutional costs	23%	7% (PASEC 2010)	49% (Uruguay, national assessment, 2003)
Test fees	16%	5% (PISA Chile, 2009)	35% (PISA Uruguay, 2003)
Other	3%	1% (PISA Peru, 2009)	7% (PISA Chile, 2009)

Note. Above calculations based on data from 13 assessments (see Table 1 for costs included in each category and for each assessment).

Table two: Costs by category, as percentage of total assessment expenditures

Appendix A

Cost categories of the assessments used in selected studies. (Adapted from Wolff (2007)).

1. Test preparation
   a. Creation and editing of test items
   b. Pilot testing
   c. Training

2. Test application
   a. Test design and editing
   b. Test printing
   c. Printing of other materials
   d. Distribution to examiners
   e. Field testing
   f. Control and supervision

3. Processing and analysis
   a. Coding and digital input
   b. Marking open-ended questions
   c. Additional analysis

4. Dissemination
   a. Report to each school
   b. Report production and distribution
   c. Public relations retainer

5. Institutional costs
   a. Personnel- in project budget
   b. Personnel- contributed *(e.g., consultants)*
   c. Infrastructure- in project budget *(physical space for personnel)*
   d. Infrastructure- contributed
   e. Equipment- in project budget *(e.g., computers and related testing equipment)*
   f. Equipment- contributed
   g. Other (*e.g., telecommunications, electricity and office supplies)*
   h. Test fees

6. Cost breakdown

a. Cost of testing per student
b. Cost of educating a student (at test-specific grade level)
c. Cost of testing as % of total budget for one grade
d. Cost of testing as % of total secondary education budget

Appendix B

Cost studies of selected national, regional and cross-national assessments

- National assessments:
  - SIMCE/LLECE 2004
  - Uruguay national assessment 2002
  - Honduras national assessment 2002

- Regional assessments:
  - SACMEQ II
    - Swaziland 2006
    - Tanzania 2006
    - Zambia 2006
  - PASEC 2010

- International assessments:
  - PISA
    - PISA Chile 2009
    - PISA Mexico 2009
    - PISA Panama 2009
    - PISA Peru 2000
    - PISA Peru 2009
    - PISA Uruguay 2003
  - PIRLS

- Hybrid assessments:
  - EGRA
    - Liberia 2008
    - Nicaragua 2008

References

Banerji, R. & Wadhwa, W. (2006). Between outlays and outcomes. *Times of India*. Accessed on February 28, 2011 at http://images2.asercentre.org/Resources/Articles/Between__outlays_-_outcomes.pdf.

Braun, H. & Kanjee, A. (2006). Using assessment to improve education in developing nations. In Cohen, J. E., Bloom, D. E., & Malin, M. (Eds). *Improving education through assessment, innovation, and evaluation* (pp. 1-46). Cambridge, MA: American Academy of Arts and Sciences.

Coombs, P. H. & Hallak, J. (1972). *Managing educational costs.* New York: Oxford University Press.

Gilmore, A. (2005). *The impact of PIRLS (2001) and TIMSS (2003) in low- and middle-income countries: An evaluation of the value of World Bank support for international surveys of reading literacy (PIRLS) and mathematics and science (TIMSS).* New Zealand: IEA.

Greaney, V. & Kellaghan, T. (2008). *Assessing national achievement levels in education*. (National assessment of educational achievement, Vol. 1). World Bank: Washington.

Grin, F. (2005). The economics of language policy implementation: Identifying and measuring costs. In N. Alexander (Ed.), *Mother tongue-based bilingual education in Southern Africa: The dynamics of implementation.* Proceedings of a Symposium held at the University of Cape Town, 16-19 October 2003, Cape Town: Volkswagen Foundation & PRAESA.

Hanushek, E. & Woessmann, L. (2005). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries (NBER Working Paper 11124). Cambridge, MA: National Bureau of Economic Research.

Heugh, K. (2006). Cost implications of the provision of mother tongue and strong bilingual models of education in Africa. In H. Alidou, A. Boly, B. Brock-Utne, Y. Diallo, K. Heugh & H. Wolff (Eds.), *Optimizing learning and education in africa –the language factor: A stock- taking research on mother tongue and bilingual education in Sub-Saharan Africa* (pp. 56-84). Paris: IIEP/ADEA.

Hoxby, C. (2002). The cost of accountability. (NBER Working Paper 8855). Cambridge, MA: National Bureau of Economic Research.

ILI/UNESCO. (1998). *Literacy assessment for out-of-school youth and adults*. (ILI/UNESCO Technical Report from Expert Seminar, Paris, June 1998). Philadelphia: International Literacy Institute, University of Pennsylvania.

Ilon, L. (1992). A framework for costing tests in Third World countries (PHREE/92/65). Washington, DC: World Bank.

Ilon, L. (1996). Considerations for costing national assessments. In P. Murphy et al. (Eds.) *National assessments: Testing the system* (pp. 69-88). Washington, DC: World Bank.
Johnson, S. (1999). International Association for the Evaluation of Educational Achievement science assessment in developing countries. *Assessment in Education, 6* (1), 57–73.

Kamens, D. H. & McNeely, C. L. (2010). Globalization and the growth of international educational testing and national assessment. *Comparative Education Review*, *54*(1), 5-25.

Koeffler, S. (1991). *Assessment design*. Paper presented at the Seminar on Measurement/Assessment Issues. Educational Testing Services. Princeton, NJ.

Lockheed, M. (2004). The purpose of good quality education. Paper commissioned for the *Education For All Global Monitoring Report 2005, The quality imperative.* Paris: UNESCO.
Lockheed, M. & Hanushek, E. (1988). Improving educational efficiency in developing countries: What do we know? *Compare, 18*(1), 21-38.

Loxley, W. (1992). Managing international survey research. *Prospects*, *22*(3), 289-296.
Ndaruhutse, S. (2008). *Grade repetition in primary schools in Sub-Saharan Africa: An evidence base for change*. London: CfBT Education Trust.

Peyser, J. & Costrell, R. (2004). Exploring the costs of accountability. *Education Next, 4*, pp. 22-29.

Porter, A. C., & Gamoran, A. (2002). Progress and challenges for large-scale studies. In A.C. Porter & A. Gamoran (Eds), *Methodological advances in cross-national surveys of educational achievement* (pp. 3-23). Washington, DC: National Academies Press.

Postlethwaite, T. N. (2004). What do international assessment studies tell us about the quality of school systems? Background paper prepared for the *Education For All Global Monitoring Report 2005, The quality imperative.* Paris: UNESCO.

Ravela, P., Arregui, P., Valverde, G., Wolfe, R., Ferrer, G., Martínez, F., Aylwin, M. & Wolff, L. (2008). The educational assessment Latin America needs (Working Paper Series No. 40). Washington, DC: PREAL.

Research Triangle Institute (RTI). (2009). *Early Grade Reading Assessment Toolkit.* Washington, DC: RTI International.

Siniscalco, M. T. (2006). What are the national costs for a cross-national study? In K. Ross & I. J. Genevois (Eds.), *Cross-national studies of the quality of education: Planning their design and managing their impact* (pp. 185-209). Paris: UNESCO IIEP.

Szucs, T., Belisari, A. & Mantovani, L. (1997). Is preventive medical care worth the cost? *Biologicals, 25*, 247-252.

Topol, B., Olson, J., & Roeber, E. (2010). *The Cost of New Higher Quality Assessments: A Comprehensive Analysis of the Potential Costs for Future State Assessments*. Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education.

UNESCO. (2000). *Dakar framework for action*. Education For All: Meeting our collective commitments. Dakar/Paris: UNESCO.

Volante, L. (2006). An alternative for large-scale assessment in Canada. *Journal of Learning and Teaching, 4*(1), 1-14.

Wagner, D. A. (2003). Smaller, quicker, cheaper: Alternative strategies for literacy assessment in the UN Literacy Decade. *International Journal of Educational Research,* 39, 3, 293-309.

Wagner, D. A. (2010). Quality of education, comparability, and assessment choice in developing countries. *COMPARE: A Journal of Comparative and International Educatio*n, 40, 6, 741-760.

Wagner, D.A. (2011). *Smaller, quicker, cheaper: Improving learning indicators for developing countries*. (In press). Washington/Paris: Fast Track Initiative & UNESCO-IIEP.

Wolff, L. (1998). *Educational assessments in Latin America: Current progress and future challenges*. Working Paper. Washington, DC: PREAL.

Wolff. L. (2007). The costs of student assessments in Latin America (Working Paper Series No. 38). Washington, DC: PREAL.

Wolff. L. (2008). *Costs and financial implications for implementing national student assessments: Lessons for South Africa.* Unpublished Presentation, November 1.

[1] Authors' addresses: Daniel A. Wagner, International Literacy Institute, University of Pennsylvania, wagner@literacy.upenn.edu; Andrew Babson, University of Michigan, ababson@umich.edu; and Katie Maeve Murphy, University of Pennsylvania, kmaeve@gmail.com. Mailing address: Dan Wagner, Professor of Education, Graduate School of Education, University of Pennsylvania, 3700 Walnut Street, Philadelphia, PA 19104-6216. Parts of this paper are drawn from a larger manuscript dealing with learning assessments in developing countries (Wagner, 2011).

[2] See Hoxby (2002) for a review. Also see Wolff (2008), p. 14, who states that, "testing in Latin America, as well as the USA, is not a significant financial burden -- constituting 0.3 percent or lower of the total budget of the level of study tested."

[3] See, for example, Johnson (1999, p. 70): [T]he cost of active participation [in IEA studies] is high, too high for many developing countries to bear. Where developing countries have been able to finance their participation, one might wonder whether the expense of that involvement could possibly be fully justified, given what has been learned, and given alternative uses for the funds consumed. How useful is it for Thailand, South Africa and Colombia to find themselves at or near the bottom of the international rank order in science, while Korea and Japan appear at the top and European countries are scattered throughout?" Also, see Postlethwaite, 2004 (p. 17), who stated that the total burden of assessment "varies between an annual international expenditure of 200,000 US dollars for SACMEQ through about 3.6 million US dollars for PISA to about 7 million US dollars for IEA. This is without counting what the countries themselves have to pay for their own staff members working on the projects and the cost of data collections."

[4] See Lockheed, 2008, p. 10, on LICs. Topol et al. (2010) provide a recent review of the US efforts the costs of more complex assessments in the US, where it is claimed, in part, that improved technology can reduce costs of increased R&D. But since LICs are, for the time being, hampered by technological constraints, the increased costs of R&D will likley end up as further bottom line expenditures.

[5] As Braun and Kanjee (2006) note, "in educational systems that lack basic resources, decisions to fund national assessments are extremely difficult to make" (p. 24).

[6] International Association for the Evaluation of Educational Achievement (IEA). See Chromy, 2002, p. 84 for a listing of major studies; also Lockheed, 2008, p. 6.

[7] According to a survey of national policy makers (Gilmore, 2005, p. 45), World Bank funding has been a key determinant of decision-making in LSEA adoption for low- and middle-income countries.

[8] "Hybrid" refers to drawing together discrete elements of various LSEAs, national curricular assessments and tests that were initially designed as cognitive assessments of reading and other basic skills. See Wagner (2011) for an in-depth review.

[9] See also: Siniscalco, 2006; Ravela et al., 2008; Wolff, 1998.

[10] For example, such a choice occurred in South Africa when it was decided not to participate in the TIMSS, citing the overall cost in time and resources (Greaney & Kelleghan, 2008, p.75). Also on South Africa, see Braun and Kanjee (2006, p. 19).

[11] These data were acquired as part of a jointly sponsored project by the Fast Track Initiative and UNESCO-IIEP (Wagner, 2011); and we thank the various agencies and their representatives for providing these data, some of which are estimates, as indicated in Table 3. Percentages are rounded to nearest whole number.

[12] Sample sizes of international assessments compiled across countries can yield much larger population totals, and numbers of participation countries continue to increase. For example, PISA (2006) had more than 400,000 students participating from 57 countries.

[13] It should be noted that not all the data were complete for each category, or reflective of full actual costs. For example, the only available PASEC data were those projected costs for the 2010 assessments; only three sources provided *test fees* data; and several sources provided no data for the *processing and analysis* or *dissemination* categories. Further, noting the ranges above, some categories demonstrated more variability than others. For example, *processing and analysis* includes average expenditures from .02% to 24.8%, while apart from three assessments (Honduras national assessment 2004, PASEC 2010 and PISA Uruguay 2003), dissemination expenditures had a mean of 5.9%. In addition, analysis would also need to account for the hidden costs or even unspecified costs discussed above – for example, costs in the *other* category for PISA Chile 2009 was over 7%.

[14] Wolff (2008, p. 19) states: "…[L]arge samples can be expanded to censal testing at a low marginal cost, since the fixed costs of developing items and pilot testing can be amortized over a larger population."

[15]Preventive medicine highlights the need for good and timely information. Timely information can make the difference between life and death or the spread of an epidemic or its curtailment. Proactive measures cost less and help avert the worst. Preventive medicine is encouraged not only to avert illness, but also to reduce costs of diagnosing and treating that illness (Szucs, 1997). Similar arguments can be made in the education arena. For instance, absenteeism and drop-out are well-known problems in LICs, incurring huge financial and social costs. Two measures have been highly successful against these difficulties: decreased grade repetition (Ndaruhutse, 2008) and increased bilingual education (Grin, 2005; Heugh, 2006). Assessments can assist in both detecting and "diagnosing" schooling difficulties earlier – from the cognitive to the socio-behavioral – thus heading off costly student problems such as dropout. In other words, even if SQC-style diagnostic tools cannot easily determine the 'best' remediation plan of action (which may be varied and complex), the early-detection aspect will nearly inevitably be cost-effective in the long run.