

The Craft of Education Assessment: Does Participating in International and Regional Assessments Build Assessment Capacity in Developing Countries?

An Independent Evaluation of IEA's Program on the Assessment
of Student Achievement (PASA)

Marlaine E. Lockheed

March 23, 2010

Acknowledgements

The author wishes to thank the World Bank, IEA, SACMEQ, PASEC, WAEC and PREAL for providing background material and information for conducting the evaluation. In addition, thanks are due to the IEA General Assembly members, TIMSS National Research Coordinators, IEA staff, PASA partner institution staff and World Bank staff who patiently answered my questions in person and via email. Comments from Pierre Varly (CONFEMEN), Ken Ross (IIEP), Patricia Arregui (GRADE), Patrick Gonzalez (NCES), and Mulikat Bello (WAEC) on the Framework for Measuring Assessment Capacity are gratefully acknowledged. Eugene Gonzales (ETS) kindly detailed the IEA-ETS Research Institute and Vince Greany (World Bank, retired) provided information about the WBI training program in assessment.

Marlaine E. Lockheed

Marlaine E. Lockheed received her Ph.D. in international development education from Stanford University in 1972. Between 1972 and 1985, she directed research on gender, education and testing at Educational Testing Service. From 1985 until her retirement in 2004, she served the World Bank as an education sociologist, sector manager, interim director of education, and head of the World Bank Institute's evaluation group. She was appointed to the U.S. National Academy of Science's National Research Council's Board in International and Comparative Studies in Education and was elected Vice-President of the American Educational Research Association. She has taught at Harvard, Stanford, Princeton, and University of Texas and is the author of 12 books and over 150 journal articles, book chapters and technical reports.

Acronyms

ACER	Australian Council for Educational Research
CONFEMEN	Conference of Ministers of Education in Francophone Countries
DGF	Development Grant Facility
ERC	Education Research Center
ETS	Educational Testing Service
FISS	First International Science Study
GA	General Assembly
GRADE	Grupo de Analisis para el Desarrollo
IEA	International Association for the Evaluation of Educational Achievement
IIEP	International Institute for Educational Planning
IRT	Item Response Theory
LLECE	Latin American Laboratory for the Evaluation of Educational Quality
MENA	Middle East and North Africa
NCES	National Center for Education Statistics
NRC	National Research Coordinator
OECD	Organization for Economic Cooperation and Development
PASA	Program on the Assessment of Student Achievement
PASEC	Programme d'Analyse des Systems Educatifs de la Confemen
PIRLS	Progress in Reading Literacy Study
PISA	International Programme of Student Assessment
PREAL	Partnership for Educational Revitalization in the Americas
SACMEQ	Southern African Consortium for the Measurement of Education Quality
TIMSS	Trends in International Math and Science Study
UNESCO	United Nations Education, Science and Culture Organization
WAEC	West Africa Examinations Council
WBI	World Bank Institute

Table of Contents

	Page
Acknowledgements.....	ii
Executive Summary.....	v
Chapter 1. What is PASA?.....	1
Chapter 2. Has IEA Implemented PASA as Planned?.....	5
Chapter 3. PASA and Capacity Building at the National Level.....	11
Chapter 4. PASA and Capacity Building at the Regional Level.....	17
Chapter 5. A Global Partnership for Assessment: What Progress?.....	22
Chapter 6. PASA and Education Policy.....	26
Chapter 7: Conclusions and Recommendations.....	32
References.....	35
Annex A – PASA Objectives and Program Components.....	37
Annex B – Evaluation Questions and Sources of Data.....	40
Annex C – Low and Lower-Middle Income Countries Participating in PIRLS 2006 and TIMSS 2007 and Receiving Fee Support from PASA or UNDP.....	41
Annex D– Framework for Measuring National Capacity for Assessment.....	42
Annex E – Low- and Middle-Income Countries with Regular National Assessments and their Participation in International and Regional Assessments.....	44
Annex F – Institutional Representatives of Low- and Middle-Income Countries at 2007 IEA General Assembly Meeting, Hong Kong.....	45

Executive Summary

IEA's PASA Program was designed to enhance the capacity of the IEA for comparative international assessment and region-specific support, and to enhance the capacity of regional assessment institutions in Africa and Latin America for comparative regional and sub-regional educational assessment and the use of assessment information. For this, IEA was awarded three consecutive World Bank Development Grant Facility (DGF) grants of \$2.5 million, totaling \$7.5 million to implement PASA; with extensions, the grant period ran from July 1, 2005 to December 31, 2009.

This evaluation addresses five main questions: (a) Has IEA implemented PASA as planned? (b) Has PASA contributed to improving the capacity of developing countries in the area of assessment of learning? (c) Have PASA partners (regional assessment bodies) undertaken activities or developed products designed to enhance the capacity of national or regional assessment institutions? (d) Has IEA taken steps to establish a global partnership for assessment? and (e) Has PASA affected educational policy in developing countries? Data for this evaluation come from reviews of documents and individual interviews with representatives of IEA, IEA General Assembly members from developing countries, national research coordinators from developing countries, World Bank staff and PASA partners.

Implementation

IEA generally implemented PASA in accordance with its plan, despite significant DGF reductions in the size and duration of the award relative to IEA's initial proposal. PASA support enabled 19 developing countries to participate in and complete PIRLS 2006 and/or TIMSS 2007, which increased the number of such countries reporting measures of learning achievement for primary and lower secondary school age children; it also allowed another eight developing countries to participate in initial stages of these large-scale assessments. IEA's website is being used, and usage increased ten-fold, from 2006 to 2008. PASA's program in training for the MENA region was completed as planned, with participation from 14 countries and stable institutional participation, and IEA's Data Processing Center has conducted training on technical aspects of large-scale assessments. Expectations for IEA's partnership with WBI in training related to national assessments did not materialize, although experts from some institutions associated with IEA or its PASA partners were involved in WBI's assessment training. Uncertainty regarding funding and the weakness of the US\$ against the Euro created a financial hardship for IEA in its efforts to include additional developing countries in the international assessments.

Has PASA built assessment capacity?

Participation in IEA studies enhances and reinforces national capacity for assessment, through "hands on" training and participation in international assessments that meet international standards for technical quality. Interviewees were consistently positive about the effects of participating in the IEA studies in building technical capacity for large-scale assessment, for managing large research endeavors and for preparing reports for policy makers. Interviewees gave highest marks to the areas of data collection, sampling, data quality control and scoring constructed responses.

It is somewhat premature to ask about PASA's impact on assessment capacity, but PASA's approach to building capacity was similar to IEA's approach in earlier large-scale assessments, including providing training on technical issues directly related to the ongoing assessment. In

addition, PASA enabled developing countries to benefit from specialized training programs offered in Arabic, and from workshops on technical aspect of large-scale assessments. Enhanced assessment capacity is demonstrated through publication of national results in IEA international reports, participation in subsequent international assessments and regularity of national assessments. The number of developing countries reporting results from IEA international assessments has steadily increased, from 5 developing countries reporting results in the First International Science Study (FISS 1970) to 33 developing countries reporting results in TIMSS 2007. Moreover, participation in IEA studies appears related to the capacity to participate in other international studies. For example, three-quarters of the developing countries participating in the first OECD study had previously participated in at least one IEA study. And nearly two-thirds of developing countries with regular national assessments have participated in an IEA study.

Have PASA partners (regional assessment bodies) undertaken activities or developed products designed to enhance the capacity of national or regional assessment institutions?

PASA partners – PASEC, PREAL, SACMEQ and WEAC -- have generally carried out the activities for which they received PASA support, including conducting training on assessment topics in their regions, and developing materials to encourage the use of assessment results for policy. PASA's activities varied across the partners: it provided bridge funding for SACMEQ during a time of instability to enable a third round of regional assessments to be carried out, supported WAEC's training of trainers for continuous classroom assessment, enabled PASEC to conduct a detailed curriculum analysis in 15 countries and build a related item bank, and facilitated PREAL's production of documents on using assessments for policy in Latin America. In some cases, country coverage was not as broad as anticipated or speed of implementation as rapid as planned. But the support from PASA has enabled the partners to carry out training and preparation of policy documents, to conduct and/or analyze data from regional surveys, and to participate in meetings related to establishing a global partnership on assessment, all of which activities enhance capacity.

Has IEA taken steps to establish a global partnership for assessment?

Progress toward achieving a global partnership for assessment is being made, albeit slowly. IEA has taken many steps in the direction by regularly convening meetings of national assessment institutions, and PASA partners have both participated in these meetings and convened regional meetings. But competition, regional interests and weak institutional capacity in developing countries has limited the effectiveness of these efforts. On the technical side, institutions cooperate in sharing and making available to developing countries software for sampling, data entry and analysis. The international and regional partnerships that exist for carrying out specific studies are consortia rather than concrete institutions. Concrete institutions exist at the national but not global level, unlike concrete global institutions with related mandates such as UNESCO's International Institute for Educational Planning (IIEP) and Institute for Statistics (UIS). Moving in this direction, IEA and ETS have recently established a virtual institute for research, training and dissemination of results from large-scale assessments, and IEA itself, including its Hamburg-based Data Processing Center, emerges as a strong center of excellence in large-scale assessment and training for assessment.

Has PASA affected educational policy?

The impact on education policy of the two major studies carried out by IEA under the PASA umbrella – PIRLS 2006 and TIMSS 2007 – was difficult to assess in 2009. The results of PIRLS 2006 became available only in late 2007 and the results of TIMSS 2007 in December, 2008. Prior evaluations had reported a positive impact on policy as seen from the perspective of participants.

Interviews for the present evaluation underscored the positive effect of earlier assessment results, particularly on one aspect of education policy: curriculum reform. Activities of PASA partners have increased dialogue regarding curriculum reform and the discrepancy between “policy rhetoric” and actual practice in several countries. The academic literature, UNESCO’s EFA Global Monitoring Reports and professional conferences have benefitted from these studies, which may, in the long run, affect education policy.

In short, despite financial difficulties caused by exchange rate fluctuations and DGF support for fewer years (three) than had been anticipated (ten), the PASA program was implemented and appears to have enhanced technical capacity for assessment in agencies of the participating countries. PASA partners have worked to build capacity for assessment in each of their four regions.

Recommendations

- *Continued World Bank support to IEA* to offset the costs to low- and middle-income countries who participate in international assessments, particularly PIRLS 2011 and TIMSS 2011. This support could come from World Bank operations, trust funds or grants.
- *Continued training by IEA* for participants in such studies, including supplemental training (such as the MENA training) for participants in low- and middle-income countries. The need for supplemental training on assessment topics in these countries was underscored in interviews.
- *Continued encouragement of larger developing countries, such as Brazil, China and India, to participate in IEA studies*, and, in some cases, to bring their expertise to the table.
- *Support for an international institution* that both conducts assessments and trains the next generation of researchers, to build capacity for large-scale national and international assessments, using IEA’s model of “learning by doing.”
- *IEA concentration on building capacity among lower-middle income countries* rather than intensifying efforts in low-income countries. The technical requirement for participating successfully in IEA studies is often beyond the existing capacity in low income countries, while lower-middle income countries frequently have a suitable basis on which to build.
- *Continued support for regional assessment institutions to work with low-income countries* to build their assessment capacity through training and participation in regional assessments and studies.
- *Establishment of a global fund for assessment* that could provide support for all the above recommendations.

Chapter 1: What is PASA?

IEA's Program on the Assessment of Student Achievement (PASA)¹ is a response both to a global demand for the assessment and evaluation of educational outcomes and to the critical need for applying professional standards in the assessment of learning achievement. Low and middle-income countries have begun to place greater emphasis on improving educational quality, and, concurrently, on methods for assessing student learning outcomes as indicators of such quality improvement. Some countries, particularly those with interest in European Union (EU) membership, seek information regarding the performance of their students compared with international or regional benchmarks. Other countries wish to participate in international or regional assessment programs that provide a comparative basis for guiding decision-making about education system reform and improvement. International donor organizations are increasingly calling for outcome-based ("results") reporting from programs they support (World Bank 2004, 2006). These concerns call for establishing or enhancing national assessment systems, so that they can provide the information requested of them. However, achieving this goal is difficult without adequate support for capacity building and institutionalization.

Through PASA, efforts have been made to build capacity and institutionalize assessments and to provide valid and reliable information on educational achievement and other indicators of education quality. PASA's strategy for capacity building rests on the assumptions that "learning through doing" is an effective approach to professional development and that the demand for evidence-based educational policy is strengthened by the availability of valid, reliable, and internationally comparable data about learning outcomes. The World Bank's Development Grant Facility (DGF) supported PASA from July 1, 2005 to December 31, 2009.²

Program Governance through IEA

PASA was implemented through the International Association for the Evaluation of Educational Achievement (IEA), a non-profit international scientific NGO licensed in Belgium for the purpose of pedagogical research worldwide. Under Belgian Law, the IEA holds tax-exempt status on all contributions, subscriptions, and grants which serve as working capital. The IEA Secretariat is registered as a foundation³ in the Chamber of Commerce, The Hague, The Netherlands (No S41158871, December 9, 1994), with offices in Amsterdam; the IEA also operates a Data Processing Center located in Hamburg. The IEA Secretariat Netherlands holds all financial rights and obligations of the IEA and employs IEA personnel; it is audited annually by the international auditing firm of Ernst & Young, The Hague, The Netherlands.

An association of national research institutions and governmental agencies, IEA is governed by its by-laws, General Assembly and Standing Committee, and is managed by an Executive Director. The Standing Committee is made of the Chair (appointed for three years) and six members who are elected from the body of the general assembly. Two new members are elected to the SC every year, and an attempt is made to secure equal representation of genders, ethnic groups and regions of the world represented. Specific studies, such as the Trends in International Mathematics and Science Study (TIMSS) and the Progress in International Reading Literacy

¹ Also referred to as the (Global) Program on the Assessment of Educational Achievement

² The initial DGF grant letter to IEA, dated September 23, 2005 covered the period July 1, 2005 to December 31, 2006. A second grant letter, dated October 16, 2006 covered the period July 1, 2006 to December 31, 2007. A third grant letter, dated January 29, 2008 covered the period July 1, 2007 to December 31, 2008. An amendment letter from the World Bank to IEA dated February 5, 2009 extended the closing of the grant period to December 31, 2009.

³ Stichting I.E.A. Secretariat Nederland (Foundation IEA Secretariat Netherlands)

Study (PIRLS), have a management group which is constituted by the main project partners. In the case of TIMSS, this is Boston College (the study center), Statistics Canada, the Educational Testing Service, the IEA Data Processing Center and the IEA Secretariat. In the case of PIRLS this group is the same with the addition of the National Foundation for Educational Research (for England and Wales). These projects report to IEA's Technical Executive Group through the IEA Secretariat twice yearly, to the IEA Standing Committee twice yearly, and to the General Assembly once a year. The IEA Secretariat employs approximately 9 full-time staff, whereas the IEA DPC employs approximately 135 full-time staff.

IEA serves as the implementing agency for PASA; it carries out its own activities and supports capacity building for regional assessment bodies in Latin America and Africa.

Program Elements

The PASA Program was designed as a 10-year, \$50 million program to enhance the capacity of the IEA for comparative international assessment and region-specific support, and to enhance the capacity of regional assessment institutions in Africa and Latin America for comparative regional and sub-regional educational assessment and the use of assessment information. The program had two components, intended to: (a) strengthen capacity for international assessment (through capacity building and institutional development, data production and analysis, international technical support, region-specific support) and (b) strengthen capacity of regional institutions for educational assessment and evaluation (through regional capacity building in Africa and Latin America). Details are provided in Annex A. The overall objective was to help build the capacity of developing countries for assessing student learning outcomes.

PASA was not funded to the level requested. Yet, with \$7.5 million Development Grant Fund (DGF) support from the World Bank over four and one-half fiscal years (July 1, 2005-December 31, 2009), PASA enabled 19 developing countries to complete and another 8 developing countries to participate at some level in the 2006 PIRLS and/or the 2007 TIMSS. Support included training/meeting costs, data management, translation verification, data use and policy analysis seminars, website expansion, a regional seminar for assessment institutes in countries of the Middle East and North Region, and a World Bank Institute program on assessment. During the same period, PASA enabled regional assessment bodies and educational policy organizations (Partnership for Educational Revitalization in the Americas, PREAL; Southern Africa Consortium for the Measurement of Education Quality, SACMEQ; Programme d'Analyse des Systems Educatifs de la Confemen, PASEC, and West African Examinations Council, WAEC) to undertake various training activities, coordination activities, and knowledge dissemination regarding assessment for countries in Latin America, Southern Africa, and West Africa. PASA also enabled IEA to convene meetings with regional assessment bodies to discuss global partnerships.

Direct support to the IEA from PASA, net of funds directed to the four partners and net of international fees for low and middle income countries' participation in PIRLS and TIMSS, was budgeted at \$400,000, or about \$110,000 per year, for the MENA region training, strengthening the website, and an independent evaluation. Actual costs were higher than anticipated for both IEA activities (training, website development, evaluation) and for ensuring the continued participation of low- and middle-income countries in the two assessments (table 1). Thus, IEA's efforts to build its own internal capacity and to take steps to establish a global partnership for assessment were financed with over \$500,000 of non-DGF resources.

Table 1: Uses of DGF grant for PASA, by purpose, as of December 31, 2008⁴, (US\$)

	PASA Grant	Actual Expenditures	Difference
IEA activities (training, website and evaluation)	400,000	480,000	(80,000)
Low and middle-income country fees and expenses	5,530,000	5,979,051	(449,051)
Partner subventions to PASEC, PREAL, SACMEQ, WAEC	1,570,000	1,570,000	(0)
Total	7,500,000	8,029,051	(529,051)

Source: IEA

Evaluation Questions

The three previous independent evaluations of IEA programs focused on the outputs, outcomes, and impact of specific IEA assessments: TIMSS 1999, TIMSS 2003, and PIRLS 2001 (Elley 2002, Aggarwala 2004, Gilmore 2005). These evaluations found that the results of such international comparable measures of learning achievement affected national policies for improving the quality of education.⁵ Moreover, the international assessments of the IEA led to the establishment of other assessment activities, beginning in the 1980s with Educational Testing Service's assessment of math and science (La Pointe 1989); OECD's first International Programme of Student Assessment (PISA) in 2000; SACMEQ in 1999; PASEC in the early 1990s; and UNESCO's Latin American Laboratory for the Evaluation of Educational Quality (LLECE) in 1997.

Previous evaluations have not, however, highlighted institutional capacity-building and progress in establishing a global partnership in assessment, which were the objectives of the DGF grant and are the foci of the current evaluation. This evaluation addresses five main questions, the first an essential precondition of the subsequent ones:

- Has IEA implemented PASA as planned?
- Has PASA contributed to improving the capacity of developing countries in the area of assessment of learning?
- Have PASA partners (regional assessment bodies) undertaken activities or developed products designed to enhance the capacity of national or regional assessment institutions?
- Has IEA taken steps to establish a global partnership for assessment?
- Has PASA affected educational policy in developing countries?

Sources of Data for the Evaluation

Data that address the first four main evaluation questions come from four major sources: (a) archival documents from IEA, PASA partners and the World Bank (including annual audits, monthly progress reports, TIMSS training materials and participant reaction evaluations, General Assembly (GA) minutes); (b) interviews with GA representatives and 16 of 19 TIMSS 2007 National Research Coordinators (NRCs) from developing countries (table 2), representatives of

⁴ Including funds that were part of the no-cost extension.

⁵ These evaluations relied heavily on interviews with IEA officials, the national research coordinators for participating countries, representatives from the Ministries of Education for selected countries, and representatives of the World Bank.

IEA and the sub-contractors, and staff of the World Bank; (c) PIRLS 2006 and TIMSS 2007 international reports; and (d) web-site usage reports from IEA and the internet. It may be premature to address the fifth evaluation question, which was more appropriate for the original 10-year PASA time frame. However, some anecdotal evidence is available from: reports presented at GA meetings, newspaper articles and web pages, the use of prior IEA study data in the preparation of donor-supported education projects and policy research, and interviews with TIMSS 2011 NRCs carried out May 11-12, 2009. Anecdotal information about the PASA partners' impact on policy in selected countries is available for review. Details are provided in Annex B.

Table 2: Interviews with TIMSS 2007 NRCs from Low and Middle Income Countries, December 2007

Countries	Population	Number Interviewed	Percent Interviewed
Self-supported	10	5	50 %
UNDP supported	7	5	71 %
PASA supported	18	16	89 %
Total Low and Middle-Income in TIMSS 2007	34	26	76 %

Study Coordinators interviewed (Salzburg, December 10-15, 2007): Armenia, Botswana, Bulgaria, *Colombia*, Egypt, *El Salvador*, Georgia, Ghana, Hungary, Indonesia, *Iran*, Jordan, Kazakhstan, Latvia, Lebanon, Lithuania, Malaysia, Mongolia, Morocco, Palestine, Romania, Serbia, *Tunisia*, *Turkey*, Ukraine, Yemen). Developing countries in italics are self-supported.

Chapter 2: Has IEA Implemented PASA as Planned?

Before PASA could have any impact, it would need to be implemented. Thus, the first evaluation question relates to the adequacy of the program in carrying out its planned activities. Sub-questions are:

- Did the PASA program enable developing countries to participate in the 2007 TIMSS and the 2006 PIRLS international assessments?
- Has the IEA website expanded and is it being used?
- Did the PASA program conduct regional seminars for assessment institutes in countries of the Middle East and North Africa Region, and how positively were these institutes perceived by participants?
- Did IEA partner with the World Bank Institute in its program on assessment?

Developing Country Participation in the 2007 TIMSS and the 2006 PIRLS

A total of 38 low and middle-income⁶ countries completed⁷ one or more of the three assessments supported by PASA: PIRLS 2006 (grade 4), TIMSS 2007 (grade 4) and TIMSS 2008 (grade 8); a list is provided in Annex C. Eight developing countries started one or more of the assessments, but did not complete them and one developing country completed the assessment but did not report results. More countries completed the 2006 and 2007 assessments than completed the comparable earlier assessments (figure 1). While approximately the same number of low and middle-income countries completed PIRLS 2006 as completed PIRLS 2001, 14 percent more low and middle-income countries completed TIMSS 2007 (either grade level) than completed TIMSS 2003.

Nearly three-quarters of the low- and middle-income countries that completed PIRLS 2006 and/or TIMSS 2007 received donor support. Twenty countries received support from the World Bank, while another 7 countries in the Middle-East and North Africa received support from the UNDP; only 11 developing countries were entirely self-supporting. However, several “self-supporting” developing countries had not paid their invoiced fees as of the time of IEA’s annual audits and their participation was supported “pro bono” by IEA (IEA Annual Audits 2006 and 2007), and one high-income country (Slovenia) received partial support.

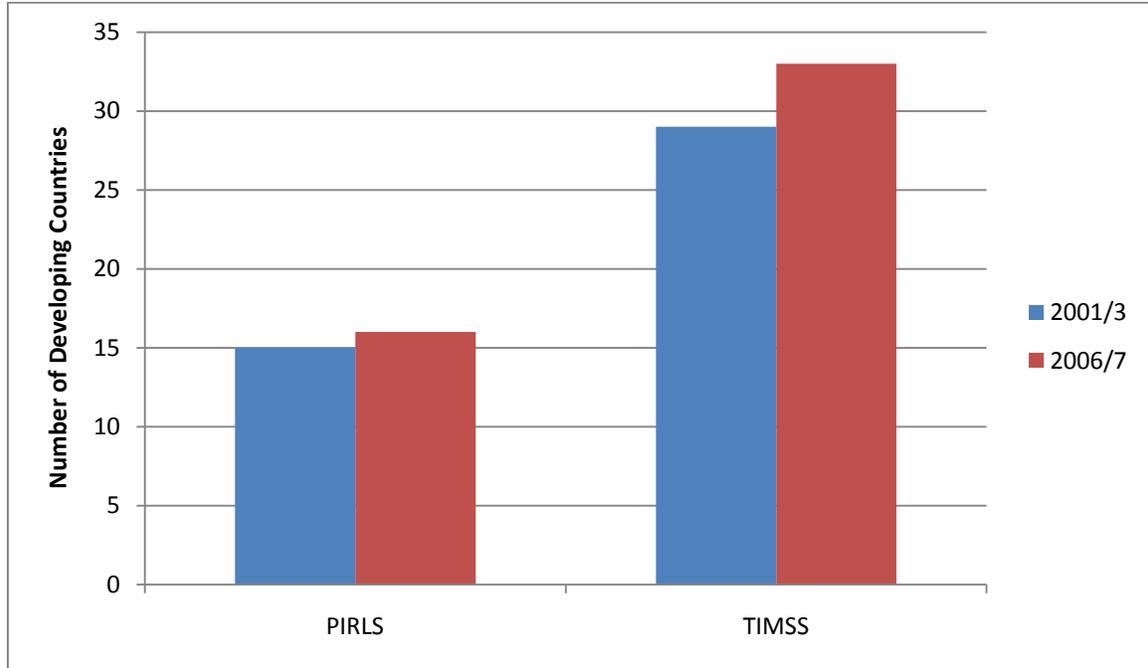
Interviews with GA members and NRCs indicated that financial support from donors was an essential precondition for their country’s participation. Specifically, 70 percent of the 26 TIMSS NRCs from developing countries who were interviewed reported that participation in TIMSS 2007 would not have been possible without the support they received through PASA or the UNDP. One interviewee added “No possibility!” and another said “Never, never ever!” in response to the question “do you think your country would have participated in TIMSS without [external support]?” This support included payment of the international fees, travel costs associated with participation in NRC meetings and local costs for data collection and entry.

⁶ As classified by the World Bank for 2005 (World Bank 2007).

⁷ Completed the study and reported data in the international report; eight countries received initial financial support through PASA but did not complete any study (Albania, Belarus, China, Honduras, Nicaragua, South Africa, Uzbekistan and Zimbabwe), one country (Moldova) received funding for both PIRLS 2006 and TIMSS 2007, but did not complete TIMSS and one country (Mongolia) completed TIMSS 2007 but the results are not published in the international report..

The financial support contributed to institutional stability, and NRCs from developing countries have participated consistently over the duration of the studies. For example, of the NRCs from developing countries, 93 percent attended the 7th TIMSS 2007 NRC meeting in Austria and 82 percent attended the 8th TIMSS 2007 NRC meeting in Botswana. Seventy-nine percent of the countries were represented at both meetings, and the same individuals attended both meeting for 54 percent of the countries. Continuity of both institutions and individuals is a strong indicator of capacity (Galiani and Coralles 2006).

Figure 1: More low- and middle-income countries completed PIRLS 2006 or TIMSS 2007 than completed PIRLS 2001 or TIMSS 2003



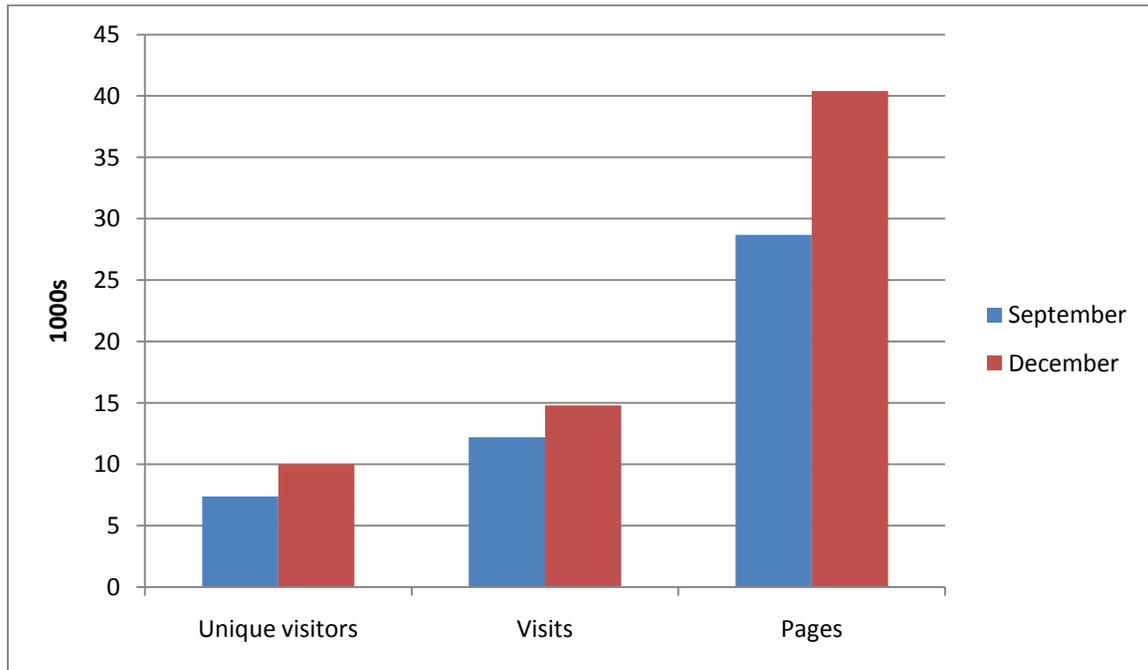
IEA Website Expansion and Use

PASA provided support to IEA to improve the “user-friendliness” of its main website (iea.nl) and incorporate a tool for on-line data analysis (iea-data.org); the cost for these improvements amounted to about \$82,000, half of which came from PASA. Usage of the main IEA website has increased significantly, 2006-2008. For example, between 23 September 2006 and 24 October 2006, the IEA website had 677 unique visitors and 745 visits searching 7524 pages; no developing countries were among the top 10 countries searching the site. For the month of September 2008, by comparison, the IEA website had over ten times the usage of 2006: 7453 unique visitors, 12,261 visits, searching 28,717 pages, and one developing country (China) was included in the list of the top 10 countries.

Developing countries increasingly accessed the IEA website. In the six months from April 10, 2008 to October 13, 2008, seven developing countries registered more than 6000 “hits” on the IEA website, averaging between 2,700 and 1000 hits per month: Mexico (15,862), Chile (15,724), Iran (11,792), China (10,793), Philippines (10,228), India (6,535) and South Africa (6,139). Another 25 developing countries registered between 1200-5999 hits on the site, for monthly averages of 200-1000 hits. This is a sharp increase from 2006, when only Mexico registered more than 200 hits in the one reference month for which data are available.

IEA website usage spiked December 9-10, 2008, following release of the TIMSS 2007 International Report on December 6, 2008. For the month as a whole, 10,082 unique visitors made 14,871 visits and searched 40,428 pages, for increases of 35 percent, 21 percent and 41 percent, respectively, relative to September 2008 (figure 2). Nine developing countries registered over 1000 hits in December: Hong Kong (2,411), China (2,026), Iran (1,814), Turkey (1,464), Mexico (1,365), India (1,356), Philippines (1,291), Malaysia (1,220), and Chile (1,133). And another 29 developing countries registered between 200 and 999 hits on the site for the month. Given that overall web site usage was minimal during the last week of the month, corresponding to the holiday period in many countries, these numbers should be adjusted upwards for comparison with other months.

Figure 2: IEA Website usage increased following release of TIMSS 2007 in December 2008



Source: IEA

Regional Seminars for Assessment Institutes in Countries of the Middle East and North Africa

PASA provided approximately \$328,000 for IEA to conduct regional seminars on assessment in Jordan and Oman in 2006, Tunisia in 2007 and Cairo in 2008 (table 3). With the exception of Djibouti, all 14 participating countries of the MENA region⁸ were represented at all seminars, which were designed to be cumulative learning experiences. Institutional representation across the four training sessions was remarkably consistent, with typical representation from research, evaluation and examination divisions of the national ministry of education.

Individual participation was also consistent for a core group of 14 individuals (19 percent of the total 75 participants) who participated in all four seminars and another 9 (12 percent) who participated in three of the four seminars. More than half of all the participants, however, attended

⁸ Algeria, Bahrain, Djibouti, Egypt, Jordan, Kuwait, Lebanon, Morocco, Oman, Palestine Authority, Saudi Arabia, Syria, Tunisia, Yemen

only one seminar (39 participants, or 52 percent); among these, 12 attended only the seminar held in their own country.

Table 3: Country participation in MENA Region PASA Training

Seminar	Date	Individuals	Topic of Training
Amman, Jordan	2/20-23/06	39	TIMSS international database, study design, subgroup scores, IEA IDB Analyzer software
Muscat, Oman	8/28-30/06	33	Analysis and report writing for policy makers
Tunis, Tunisia	2/18-22/07	37	Multiple-item indicators, framework for writing policy-relevant research
Cairo, Egypt	1/13-17/08	41	School effectiveness analysis and relevant software (HLM) for analysis

Source: IEA

Topics of the training focused on methods for data analysis and frameworks for writing policy-relevant research. Participant evaluations for the seminar in Jordan were provided for review, and showed a generally high degree of satisfaction with the various technical topics of the training. However, evaluations were submitted by only 27 of the 39 participants. Of these, 70 percent (20 participants) reported that they “Agreed a lot” with the statement that “The seminar will be helpful for my further work” and another 22 percent (6 participants) reported that they “Agreed a little” with the statement.⁹

In terms of results, two-thirds of the individuals who participated in all four training sessions completed policy research papers that have been subsequently incorporated as chapters in *Issues in the Middle East and North Africa Region* (Lietz and others 2008). These chapters document issues of equity and quality in basic education in Algeria, Egypt, Jordan, Morocco, Oman, Palestine, and Yemen. Interviews with World Bank staff working on education in the MENA region confirmed the effectiveness and utility of the seminars, which were carried out in Arabic, in creating a regional network of expertise.

IEA Partnership with the World Bank Institute

IEA did not directly partner with the World Bank Institute (WBI) during the project period. During FY06-FY08, the World Bank Institute offered courses dealing with assessment in Uganda (February 2, 2006), Kuwait (March 17-28, 2007) and Pakistan (April 21-26, March 3-7, and May 26-30, 2008). The IEA Executive Director gave a videoconference presentation to participants of the Uganda training program, but otherwise IEA had no direct partnership for these programs.

Some of the instructional experts for the WBI programs, however, came from organizations with a long history of association with IEA, including the Australian Council for Educational Research (ACER), the Education Research Centre (ERC) of St. Patrick's College in Ireland, Educational Testing Service (ETS), and Statistics Canada (StatsCan).¹⁰ For example, the Kuwait training brought together participants from countries that had participated in earlier WBI training Programs for additional sessions on classical test theory and item response theory (IRT) and to make country presentations; experts came from WBI, the Canadian Council on Learning, the ERC and StatsCan. The Pakistan training was part of a World Bank project and involved experts from the ERC, ACER, ETS and StatsCan. In addition, the PASA partner PREAL worked with WBI on training in Latin America. Specifically, WBI and PREAL offered training on curriculum analysis for test development in the Dominican Republic (January 23-31, 2008) and on writing

⁹ The other two options were “Disagreed a little” and “Disagreed a lot.”

¹⁰ ACER and ERC are both members of the IEA General Assembly; StatsCan has a representative on the IEA Technical Committee, and ETS is affiliated with TIMSS.

reports on the use of results from national assessments in Panama (June 9-13, 2008); a third course on the topic of statistical analysis and measurement theory was held in El Salvador (March 2009). Some of the experts in these courses had prior experience with IEA international assessments.

Funding for PASA

Four issues regarding the funding of PASA were: changing the grant application from Window 1 of the DGF to Window 2, reducing the size of the grant, delayed timing of the awards, and the impact of the devaluation of the US dollar relative to the Euro on the purchasing power of the grant. The overall impact of these issues led to a total budgetary shortfall of approximately US\$1 million.

The duration of PASA was reduced by the World Bank from the 10 years requested to three years. The proposal for PASA called for a 10-year program, to be supported under “Window 1” of the World Bank’s DGF; the actual award was limited to three years, under “Window 2” of the DGF, which establishes a three-year maximum for awards. IEA expressed considerable concern with the World Bank for having shifted the proposal from the first to the second window without changing the parameters of the PASA program, and the IEA Executive Director communicated with the World Bank on this subject.

The DGF award was 85 percent lower than the original request. IEA initially requested \$5 million per year for ten years for PASA under Window 1, for a total of \$50 million, to enable developing countries to participate in the 4-5 year cycles of PIRLS 2006 and TIMSS 2007 (through payment of international fees and additional support for travel, training and incidental national costs) and to support regional partners. This award was reduced to three \$2.5 million per year grants, only 15 percent of the original request, with no reduction in the scope of work to be carried out under PASA. As a consequence, IEA incurred approximately \$500,000 in expenses for developing country participation that the DGF did not cover. Regrettably, some fees were paid for countries that later did not complete the assessments (about \$560,000) and for one high income country (\$147,000). At the same time, resources for building IEA’s capacity, training and building a global partnership for assessment amounted to about \$400,000, most of which went to the MENA training. Costs for website development and evaluation were also somewhat higher than initially budgeted.

Funds from the DGF were not awarded in a timely manner, with grant letters being sent three to seven months after the start of the World Bank’s fiscal year. The DGF awarded funding for PASA to IEA in three equal tranches of \$2.5 million, through award letters dated September 23, 2005 (for expenses July 1, 2005 to December 31, 2006), October 16, 2006 (for expenses July 1, 2006 to December 31, 2007) and January 29, 2008 (for expenses July 1, 2007 to December 31, 2008). In addition, IEA received a 12-month no-cost extension for the program to enable uncommitted funds to be committed through December 31, 2009; this extension principally affected the partners that had under-spent their awards as of end of the original grant period. The 15-month hiatus between the second award and the third award created uncertainty and some cash flow issues for IEA, which incurred out-of-pocket expenses to allow developing countries to participate in ongoing organizational and training activities for both PIRLS and TIMSS. This hiatus was not well communicated to the PASA partners, which also experienced some difficulties as a consequence. In addition, the hiatus created some reputational risks for the DGF.

Finally, as pointed out by the Executive Director in his letter to the World Bank Managing Director responsible for the DGF, the purchasing power of the US\$ relative to the Euro declined significantly from 2005 to 2009. This created a hardship for IEA, since approximately 35 percent

of the program costs were expended in Euros. IEA estimated the overall impact to be about \$488,000.

Conclusion

IEA generally implemented PASA in accordance with its plan, despite substantial DGF reductions in the size and duration of the award relative to IEA's initial proposal. PASA support enabled 38 developing countries to participate in and complete PIRLS 2006 and TIMSS 2007, which increased the number of such countries reporting measures of learning achievement for primary and lower secondary school age children. IEA's website is being used, and usage increased substantially, from 2006 to 2008. IEA's program in training for the MENA region was completed as planned, with participation from 14 countries and stable institutional participation. Expectations for IEA's partnership with WBI in training related to national assessments did not materialize, although experts from some institutions associated with IEA were involved in the training, but other training on technical aspects of large-scale assessment was offered through IEA's Data Processing Center. Uncertainty regarding funding and the weakness of the US\$ against the Euro created a financial hardship for IEA in its efforts to include developing countries in the international assessments.

Chapter 3: PASA and Capacity Building at the National Level

Capacity building is a long-term process, and short-term results may elude measurement. We define “assessment capacity” along four dimensions: (a) technical capacity for developing the assessment and accompanying survey questionnaires, (b) administrative capacity regarding the physical (or electronic) preparation of tests and related materials including any necessary translation, test administration, professional scoring of answers to test questions, and data entry including quality control, (c) analytic capacity regarding sampling, test reliability, scale score construction, analysis and (d) communication capacity for dissemination of results (see Annex D for details in each of these areas). Institutional capacity for assessment, which Galiani and Corrales (2006) identify as stability, coherence and coordination and quality of implementation, is also important. The PASA program, proposed for a 10-year period of implementation, was designed to build institutional capacity, but the actual 4.5-year implementation period was not sufficient to achieve this objective comprehensively; the definition of “capacity” used in this report would fall under Galiani and Corrales’ dimension of “quality of implementation.”

The PASA program, which concentrates on “learning by doing” assessments, may have enhanced national technical, administrative and communication capacity, while raising awareness regarding analytic capacity needs. Building analytic capacity, particularly in the narrow fields of sampling and scale score construction, has been less of a focus for PASA, since expertise in these areas is generally concentrated in only a few centers of excellence, worldwide. Instead, PASA has provided turn-key software and technical assistance in these areas.

The process of carrying out an IEA study involves a number of concrete technical steps, for which the IEA offered training to representatives from participating countries (see example in table 4). In addition, for countries in the Middle East and North Africa (MENA) region, specialized workshops were held on secondary analysis of TIMSS data for writing technical and policy reports.

Table 4: IEA training for TIMSS 2007 covered a range of implementation topics

Date	Locale	Training Activity	N
2/7-11/2005	Egypt	First NRC meeting for TIMSS 2007 (in conjunction with TIMSS 2003 International Database training)	58
5/30-6/3/2005	Netherlands	Reviewing draft assessment framework, writing mathematics and science items for the field test	62
11/13-18/2005	Slovenia	Reviewing field-test instruments	62
12/13-18/2005	Germany	Field-test data management seminar	
3/5-10/2006	Malta	Training for scoring constructed-response items for the field test	61
8/6-11/2006	Latvia	Reviewing field-test results and finalizing achievement tests and questionnaires for the main data collection	58
9/18-21/2006	Germany	Data management seminar	
3/4-9/2007	Jordan	Training for scoring constructed-response items for the main data collection	50
10/11/2007	Dubai	Data management seminar	
12/9-14/2007	Austria	Reviewing draft exhibits for the TIMSS 2007 International Reports, reviewing country data with technical specialists	60+
6/22-27/2008	Botswana	Reviewing draft text for TIMSS 2007 International Reports, reviewing exhibits for the mathematics and science reports, reviewing exhibits for the TIMSS 2007 Encyclopedia, training in analysis of national data	44
2/8-13/2009	Egypt	Training on the use of TIMSS 2007 International Database.	

Source: IEA TIMSS 2007 Progress Reports

Has “learning by doing” been effective in building assessment capacity, and what would indicate success? Three indications of enhanced assessment capacity are: (a) the self-reports of study participants, (b) the regularity of a country’s participation in international assessments and (c) the regularity and quality of its own national assessments.

Participant Self-Reports

TIMSS National Research Coordinators (NRCs) were consistently positive about the effects of participation in the IEA studies and related training in building technical capacity for assessment, for managing large research endeavors and for preparing reports for policy makers. For example, in response to the question “What specific skills or knowledge have you learned or expanded through your participation in TIMSS”, PASA country participants noted:

- “We had never done this type of test before. Every aspect of testing I’ve learned in TIMSS. All logistics, training teachers and test administrators, scoring open questions, all computer-related questions, use of secure plastic bags to distribute tests.”
- “Main thing was principles of sampling; I got familiar with different types of procedures for sampling. In questionnaire development, in translating and revising. In data quality control and the use of programs to ensure quality.”
- “Research methods, including everything!”
- “Learned a lot: management as a NRC, to manage the whole process. How to build a national data base, enter data, control quality, administer tests, how to score items, how to write manuals, how to work in a team. If I had read tens of books I would not have learned what I learned from participation in TIMSS.”
- “Learned a lot on project management – how to manage a large research project. The tasks and deadlines were well defined, which was different from our normal situation, in which we end up doing things at the last minute.”
- “How to use data management software (WIN W35, WINDEM), how to organize ‘such a huge assessment’ (contacting schools and principals through phone, email, letters with stamp of Ministry of Education), adapting questionnaires to national needs, many things.”
- “Analysis—how to use data and make reports for policy makers.”
- “Policy report writing based on empirical evidence; we used to write reports based on expert opinion only.”

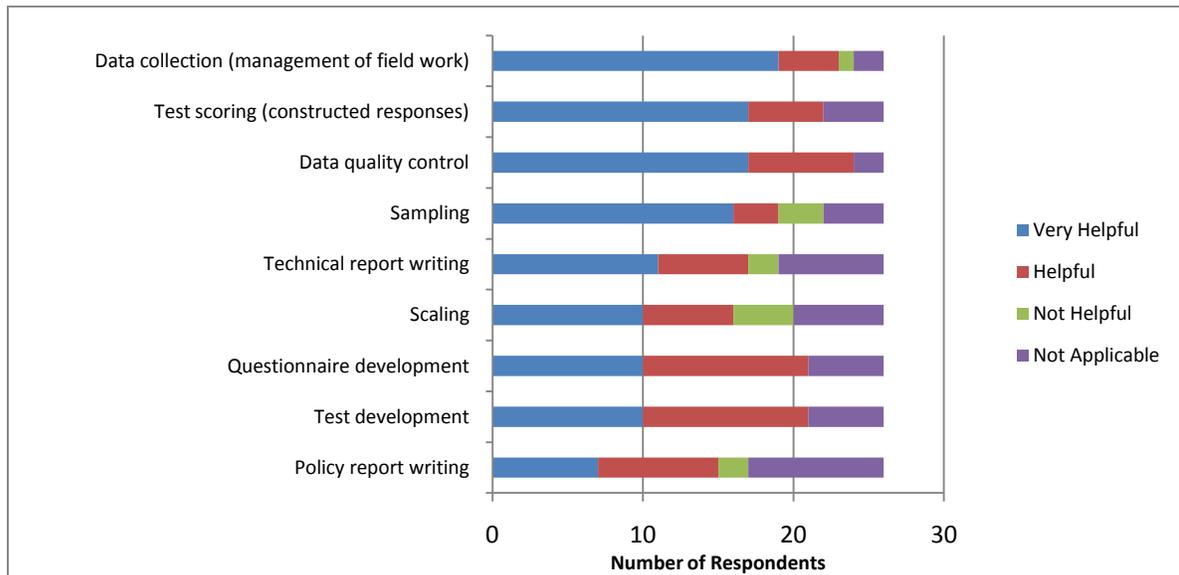
Moreover, in some countries, the capacity building had spillover effects on national assessments:

- “Our national assessment learned from the international assessment.”
- “Helped in writing national assessment items, how to interpret data, using IEA software for data entry of national assessment, disseminated IEA items to teachers to improve their question-writing skills, became aware of IRT.”

In response to the question “How would you rate the special or on-the-job training that you and your TIMSS team received during TIMSS” on a scale of “Very helpful”, “Helpful”, “Not helpful” or “Not applicable”, NRC interviewees gave highest marks to the areas of data collection, sampling, data quality control and scoring constructed responses (figure 3). Since the interviews

were carried out in advance of the policy and technical writing activities of TIMSS 2007, only those interviewees who had participated in either the specialized writing workshops for the MENA region or in previous TIMSS assessments answered the questions about these two topics.

Figure 3: Training related to the management of large scale assessments was rated most highly by NRCs from developing countries (N = 26).



Source: Author's interview data

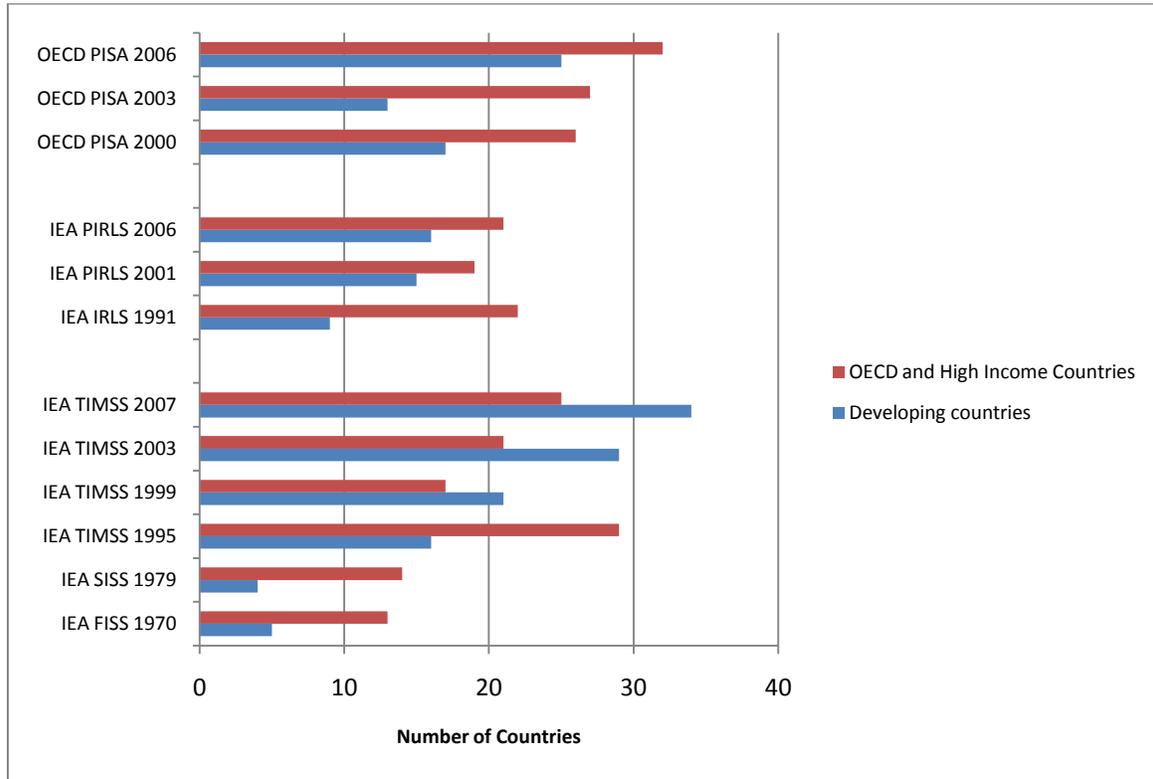
When asked to assess their country's current capacity for carrying out student assessments, respondents attributed higher capacity in the areas of project coordination, technical activities closely related to in-country activities of TIMSS (sampling, test development, questionnaire development, data collection, and data entry and quality control) and in-country communication about the assessment. They attributed lower capacity to technical areas for which the IEA provided software and technical support: scoring, scaling, technical report writing and policy report writing.

Developing Country Participation in International Assessments, 1970-2007

Participation in an IEA study appears to lead to participation in subsequent international assessments. Since 1970, the number of developing countries¹¹ participating in international assessments sponsored by IEA has increased, from 5 countries completing the First International Science Study to 34 completing TIMSS 2007. Since 1999, more developing countries than developed countries have participated in TIMSS, reaching a high of 58 percent of participating countries in 2007 (figure 4).

¹¹ Developing countries classified by the World Bank (most recently, in 2005) as Low Income, Lower Middle-Income and Upper Middle-Income that have participated in one or more international assessments are: Albania, Algeria, Argentina, Armenia, Azerbaijan, Belize, Bosnia and Herzegovina, Botswana, Brazil, Bulgaria, Chile, Colombia, Croatia, Czech Republic, Egypt, El Salvador, Estonia, Georgia, Ghana, Hungary, India, Indonesia, Iran, Islamic Rep. of, Jordan, Kazakhstan, Kyrgyz Republic, Latvia, Lebanon, Lithuania, Macedonia, Rep., Malaysia, Mexico, Moldova, Mongolia, Morocco, Nigeria, Oman, Palestinian National Authority, Peru, Philippines, Poland, Romania, Russian Federation, Serbia and Montenegro, Slovak Republic, South Africa, Swaziland, Syrian Arab Republic, Thailand, Trinidad and Tobago, Tunisia, Turkey, Ukraine, Uruguay, Venezuela, Yemen, Zimbabwe.

Figure 4: High income and developing countries participating in PISA, PIRLS and TIMSS, 2000-2007



Source: Author's analysis of IEA and OECD data

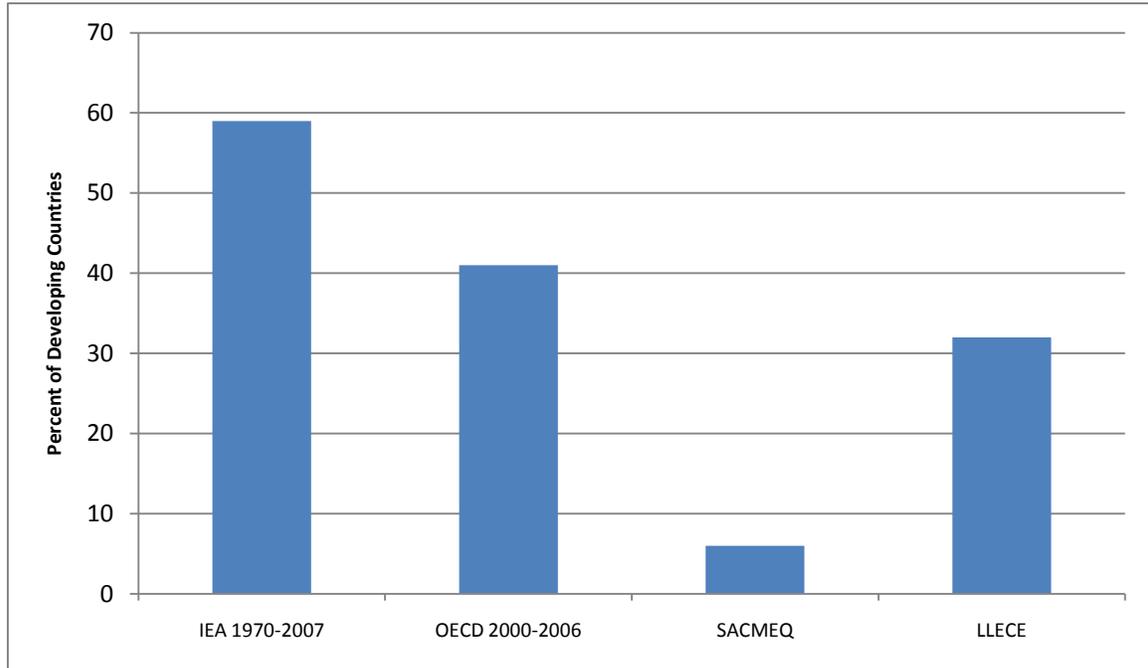
Moreover, prior participation in an IEA study seems related to participation in subsequent studies under different sponsorship. For example, of the 17 developing countries participating in the first OECD study, PISA 2000, 13 (76 percent) had previously participated in at least one IEA study. And 68 percent of the 25 developing countries participating in the most recent OECD study, PISA 2006, had participated in an earlier IEA study.

Participation in IEA Studies and National Assessments

Participation in international assessments also appears to be related to regularity in national assessments. The total number of countries that have carried out at least one “national assessment” (including census-based and high stakes assessments) grew from 65 countries in 1995-1999 to 111 countries in 2000-2006, including 80 developing countries (Benevot and Tanner 2007). Forty-one percent of these 80 developing countries have participated in at least one IEA study, with significant regional variation: fewer than 25 percent of countries with recent national assessments in sub-Saharan Africa have participated in an IEA study, whereas 75 percent of Arab states with national assessments have done so (see Annex E).

Fewer (34) developing countries carry out annual or regularly periodic national assessments. However, nearly two-thirds of developing countries that have done so have also participated in one or more IEA studies, a higher share than have participated in other international or regional assessments (figure 5).

Figure 5: A high share of developing countries with annual or regularly periodic national assessments have participated in IEA studies



Note: No country with annual or regular periodic national assessments participates in PASEC.
Source: Author's analysis (see Annex E).

Indicators of the quality of these national assessments are generally unavailable. Two comprehensive studies in Latin America review institutional capacity for assessment, but say little about the quality of the assessment instruments themselves (Galiani and Corrales, 2006; Ferrer 2006). For example, Ferrer notes: “Technical validation of test items or questions is a critical element in developing assessment instruments; unfortunately, national technical reporting on this subject is not particularly detailed, so it is difficult to provide much information about technical validation experiences in Latin America” (Ferrer 2006: 28). Lockheed (2008) observes that “little public information is available about the technical characteristics of national learning assessments” in developing countries (Lockheed 2008:6).

Alignment of Institutional Capacity Development with Relevant Institutions

The PASA program offered a variety of activities targeted at research and evaluation institutions, only some of which are involved in national assessments. These activities include the training for PIRLS and TIMMS, the MENA training, and the IEA General Assembly meetings.

Considerable alignment between those participating as NRCs for TIMSS 2007 and those responsible for national assessments was observed. Interviews with 25 NRCs from developing countries at the TIMSS 2007 NRC meeting in December 2007 revealed that 60 percent of the participating developing countries were represented by an institution or individual that was directly responsible for other national assessment activities (examinations, national assessments, test development) and another 10 percent were responsible for providing advice on assessment.

Several commented that participation in the TIMSS had spill-over effects on the national assessments.

Less alignment was observed for participants at the 2007 GA meeting. A review of the participation lists from the 2007 GA meeting showed that, of the 29 low and middle income countries, about one-third (11) of the representatives came from organizations that also were responsible for a national learning assessment, according to UNESCO. The remaining two-thirds were equally divided between representatives from countries lacking a national assessment, where participation in the IEA studies could encourage these countries to undertake national assessments, and representatives from apparently different organizations from those responsible for national assessments (see Table 5 and Annex F). In all cases, GA members represent organizations that are authorized by their governments to implement the IEA studies.

Table 5: Non-alignment between IEA GA member institutions and national institutions for assessment

Low and Middle Income countries at 2007 IEA GA meeting	Institutions represented at GA (Responsible for IEA studies)	Institutions responsible for National Learning Assessment (per UNESCO)
Bulgaria	University of Sofia, Faculty of Mathematics and Informatics	Ministry of Education
Estonia	Tallinn University	National Examination and Qualification Center
Hungary	Ministry of Education	CES, OKEV, CADR
Indonesia	MOE National Office for Educational Research and Development	Educational National Standards Board
Jordan	National Center for Human Resources Development	Ministry of Education
Malaysia	Ministry of Education, Educational Planning and Research Division	Ministry of Education, Malaysian Examination Syndicate
Philippines	Science Education Institute, Department of Science and Technology	Ministry of Education, National Education Testing and Research Center
Romania	Institute of Educational Sciences	National Assessment and Examination Service
Thailand	Office of the Education Council	National Institute of Education Testing Service

Source: IEA (2007) and UNESCO (2007)

Conclusion

Direct training and participation in various IEA activities has contributed to improved capacity for assessments, as indicated by participant self-reports, regularity of participation in international assessments and regularity of national assessments. Some alignment between those participating in the IEA studies and those responsible for national assessments was observed, and participation in the IEA studies may have contributed to building a culture of assessment. The lack of information on the actual quality of national assessments (in contrast to the availability of substantial technical details on the quality of international assessments) means that it is not possible to determine the impact of PASA on the quality of national assessments.

Chapter 4: PASA and Capacity Building at the Regional Level

The four PASA partners undertook various approaches for building capacity at the regional level. PASEC and SACMEQ focused on building technical capacity for implementing national and regional assessments; PREAL focused on building capacity for using assessment data in policy dialogue; WAEC concentrated on training of teachers and testing professionals. This chapter discusses each of the partners separately.

Programme d'Analyse des Systems Educatifs de la Confemen (PASEC)

PASEC is a program of the Conference of Ministers of Education in Francophone Countries (CONFEMEN). It was launched in 1991 with the aim of providing decision-makers in developing member countries with objective information on their basic education systems to facilitate the development of efficient education policies. PASEC conducts quality assessment in primary education, trains national teams for carrying out assessments, and diffuses results as well as assessment and analytical tools and models for use at the national, regional and global levels. CONFEMEN counts 41 member countries, of which 25 are located in Sub-Saharan Africa and the Indian Ocean, 17 of which have already benefited from an assessment, with another 4 countries having joined in 2009. None of the PASEC countries regularly conduct national assessments, and the PASEC activities have introduced the concepts of assessment in these countries (see Annex E for list of countries with national assessments).

PASA support was used to: (a) upgrade PASEC's technical capacity for assessment to comply with the international standards prescribed for comparison of results between countries; (b) initiate partnerships with other regional assessment programs to ensure better comparability of results; (c) conduct training on aspects of assessment. Support for the latter was for the development of training material, organization of training sessions, and technical assistance (co-ordination and training).

Major activities carried out under the PASA program were (a) analyses of the curricula of fifteen PASEC countries, (b) preparation of an item data bank, and (c) training workshops in data analysis of PASEC tests and questionnaires. According to the PASEC final reports for 2006 (dated January 2007) and 2007 (dated October 2008), the program achieved 80 percent of the first activity, 100 percent of the second activity and delivered two of the three planned workshops. Specifically, for the first activity, the official curricula from 15 countries were analyzed, surveys of teachers regarding the implemented curricula were carried out in five countries, guidelines for improving PASEC tests were developed and a synthesis document on the curricular analysis was completed. However, little progress was made in establishing minimum level of competencies in curricular areas or in establishing a francophone group of test development experts. For the second activity, a database of test items was developed from national tests results in seven countries. For the third activity, two training workshops on the analysis of survey data were held in Dakar and Cotonou (November 27-December 8, 2006). In addition, PASA enabled PASEC to publish and disseminate its data on its own website.

In terms of capacity building, the PASEC report notes that "Such curricula analysis using quantitative methods and international nomenclature had never been undertaken in francophone Africa before." PASEC also notes that the PASA program was instrumental in expanding PASEC's contacts with assessment experts from IEA, SACMEQ, Research Triangle Institute (for the Early Grade Reading Assessments, EGRA) for purposes of exchange of information. PASA also enabled PASEC to establish and maintain contact with experts from the other PASA partners

and built its own capacity for assessment (for example, to change its methodology of sampling for replacement schools).

Southern African Consortium for the Measurement of Education Quality (SACMEQ)

SACMEQ is closely associated with UNESCO's International Institute of Educational Planning (IIEP). Beginning in the early 1990s, IIEP offered training programs based on IEA studies of education quality (Ross and Genevois, 2006). Participants in these programs carried out class exercises that led to SACMEQ I in 7 countries in sub-Saharan Africa (1995-97), SACMEQ II in 14 sub-Saharan African countries (2000-2002) and subsequently the formation by a group of African education ministers of an intergovernmental agency called SACMEQ, based in Zimbabwe, and governed by the ministers of education of participating countries. PASA supported SACMEQ III (2007) and data have been collected, with reports planned that are modeled after the international reports of IEA.

PASA was intended to support SACMEQ in building assessment capacity in sub-Saharan Africa, but over the course of the grant period, financial flow difficulties and a deteriorating situation in Zimbabwe initially delayed the work. Specifically, the SACMEQ Director left to take a position with UNICEF in Kenya, the Acting Director and the Data Processing Manager transferred to Paris and the offices of SACMEQ temporarily relocated to IIEP. After this, funds were used to support activities of SACMEQ III, but most expenditures were made in Paris at IIEP to ensure continuity of SACMEQ staff. It is expected that the SACMEQ offices will be relocated in sub-Saharan Africa in 2010.

During the period of the grant, PASA funds covered expenditures for: the salaries for the Acting Director and Data Processing manager (\$165,000), their travel (\$12,000), computers to ensure security and confidentiality of SACMEQ data (\$20,000), website server upgrade (\$15,000), data processing for pilot testing phase (\$25,000), preparation of main survey data collection instruments by IIEP staff (\$25,000), support for sampling @ \$2,000/country for 15 countries (\$30,000), and a printing subsidy to countries for production of instruments (\$30,000). It was anticipated that the purchased computers would be transferred to SACMEQ offices in sub-Saharan Africa in 2008-9, but a SACMEQ office in Africa has not yet been established. Despite some initial delays, as of September 2008, 12 of the 15 participating countries had made substantial progress in cleaning the final survey data. A "no-cost" extension of the PASA for the calendar year 2009 allowed SACMEQ to utilize funds that had been unspent due to its disruptions; again, these funds allowed for continuity of SACMEQ staff (\$40,000), IIEP coordination staff (\$90,000) and administrative overheads (approximately \$17,000).

Because SACMEQ is an association of education ministers, its studies translate their policy concerns into instrument design and data collection. The SACMEQ reports are organized around the initial policy concerns that motivated the study development. In carrying out the studies, SACMEQ has helped build capacity in sampling through use of the IIEP sampling software (IIEPSAMP), familiarized participating country institutions with fundamental standards of test construction based on curriculum and textbooks (classical test construction), introduced Rausch Item Response Theory (IRT) scaling (necessary for looking at change over time and vertical scaling) for the more technically advanced countries, and provided extensive training in data collecting and cleaning. A major contribution of PASA to SACMEQ was IEA's WIN Data Entry Manager (WINDEM), used for teaching and data entry of SACMEQ III data; as of September 2008, over half of the 15 SACMEQ III countries had completed over 50 percent of their data cleaning.

Partnership for Educational Revitalization in the Americas (PREAL)

For the LAC region, PASA supports PREAL in its efforts to enhance the effective use of education assessment data collected from national assessment initiatives. PREAL is a region-wide network of institutions seeking to improve education policy. It was established in 1995 by the Inter-American Dialogue (in Washington) and CINDA (in Santiago, Chile) as a multi-year initiative aimed at building a broad and active constituency for education reform in many countries. It has become the leading non-governmental voice on education in Latin America and a strong advocate for involving leaders from civil society in the work of education reform. A majority of PREAL's activities are carried out in collaboration with expert public policy centers throughout the region working to promote education reform.

One of the major weaknesses in assessment systems is the failure to use assessment information to guide policies, programs and investments decisions. PASA provided PREAL support to work with more than 15 Latin American countries that have completed national assessments under World Bank financed projects to carry out activities designed to increase the knowledge base on assessment and improve its dissemination and use, including the following:

- Hold conferences and seminars to expand the region-wide network of specialists and professionals for a more active involvement in educational assessment and the use of assessment data to conduct studies of policy relevance to LAC.
- Prepare reports and publications on good practices in assessment.
- Expand the on-line clearinghouse on assessment to include translated information from all over the world.
- Adapt and translate existing training materials to meet regional demands and contexts.
- In partnership with WBI, offer training in assessment to specialists in Latin America; in 2008, specialists from 13 countries attended training on curriculum analysis and instrument design in the Dominican Republic and training on report writing in Panama.

PREAL completed numerous publications with support from PASA, including two working papers (*The Costs of Educational Assessments in Latin America* and *The Educational Assessments that Latin America Needs*), one study (*A Comparative Study on the Development and Implementation of Curricular Standards in Latin America*), a manual on education standards (*Standards in Education: Implications for their Application in Latin America*) and numerous newsletters, policy briefs and one-page summaries related to assessment issues, typically made available in both Spanish and English. PREAL reported that ministries of education in Honduras, Colombia, Ecuador and Guatemala had requested various reports and publications, and concluded that this represented high-level government interest in PREAL and its publications. Publications were also requested by NGOs: the Federation of Private Development Organizations and the Association of Education and the Family.

In cooperation with the World Bank Institute, PREAL developed and delivered four training workshops in 2008 and 2009: (a) for journalists related to the dissemination of assessment results for government officials and media professionals, (b) for ministry officials and technical specialists on curricular analysis and test design, (c) for policy specialists on dissemination and utilization of test results, and (d) for technical specialists on classical item analysis and IRT analysis. Standard WBI Level-1 evaluations of these workshops were not available for review, but reports received from PREAL suggest that ministry officials who attended were satisfied with the training and that it influenced specific technical and policy decisions that they were considering.

PREAL also contributed to the development of a network of specialists working on issues of assessment, through its Working Group on Standards and Accountability, the Ibero-American Network for Research on Education Change and Efficacy, and UNESCO's International Bureau of Education. The Working Groups has been involved in studies of in-classroom student assessment practices and the use of assessments for pedagogical purposes, and various replications of these studies are underway.

West Africa Examinations Council (WAEC)

WAEC's focus under PASA was training and staff development for improved testing. The major WAEC activities supported under the PASA program were (a) conducting a needs assessment for and training teachers for continuous assessment, (b) training examiners, item writers and moderators for the West African Senior Secondary Certificate Examinations (WASSCE), and (c) training test development and research staff of WAEC. While the initial proposal focused on all five member countries of WAEC (Ghana, Nigeria, The Gambia, Liberia, and Sierra Leone), financial constraints reduced participation to the first three only. Interim and final reports were received that document WAEC's use of the DGF funds. The activities supported by the DGF cost an approximate \$153,700, of which the WAEC contributed approximately \$54,000 and ministries in Ghana and The Gambia contributed approximately \$10,000.

According to the final report presented to the IEA as of July 2008, WAEC completed the needs assessment and training to improve the validity and reliability of school-based continuous assessment for all three countries, training 145 trainers from Ghana, 63 trainers from Nigeria and 30 trainers from the Gambia. The participants expressed satisfaction with the training, noting that "trainers left the workshop with confidence and were full of assurance to WAEC that they were ready to roll out the program" in their respective locales, if funding became available. WAEC reported that some trainers in Ghana had started implementing the program for training teachers. The cost for this activity was approximately \$44,000, to which the DGF contributed approximately \$24,000 (55 percent).

The report indicated that the training for examiners, item writers and moderators for the WASSCE was completed in Ghana, planned for The Gambia (in October, 2008) but had not taken place for Nigeria. According to the report, the training for Ghana cost approximately \$56,600, with \$24,000 (44 percent) contributed from the DGF, while the training planned for The Gambia would cost approximately \$17,500, with approximately \$15,000 (85 percent) contributed from the DGF.

The training for test development and research staff of WAEC took place in Accra and involved staff from all three countries: 16 from Ghana, 6 from The Gambia and 22 from Nigeria. According to the reports, "participants indicated that they found the workshop very revealing and fruitful and that they benefitted a great deal professionally from the program." It is not possible to determine from the reports whether, as a consequence of these trainings, subsequent versions of the WASSCE (December 2007, May 2008 or December 2008) or other WAEC examinations exhibited higher quality than earlier versions. However, since the time line for test development is typically longer than a year, it is unlikely that these training activities would have improved tests offered in late 2007 or 2008. The cost of this training was approximately \$30,000, to which the DGF contributed approximately \$19,000 (64 percent).

Conclusion

PASA's impact on regional assessment and policy analysis bodies has varied, depending on the objective of the organization. For SACMEQ, PASA provided financial continuity at a time of disruption, which enabled SACMEQ III to proceed, approximately on schedule, despite the lack of an institutional base in Sub-Saharan Africa. For PASEC, PASA supported essential activities that served as a foundation for regional assessments and broke new ground. For WAEC, PASA support enabled the training to improve continuous assessment in schools and strengthen technical skills for staff working on a secondary school leaving examination. For PREAL, PASA provided funds for training, research and publications to inform dialogue on education policy, which were utilized throughout the region by ministries of education and NGOs working in education. Overall, PASA supported activities that built assessment capacity incrementally as appropriate to the initial situation in the four regions.

Chapter 5: A Global Partnership for Assessment: What Progress?

PASA was intended to build a global partnership for assessment, principally through IEA's annual General Assembly meetings, which bring together representatives from developed and developing countries and international agencies to discuss topics related to learning assessment. For example, in 2007, GA representatives from 41 IEA member countries, the four PASA sub-contracting institutions (PREAL, WAEC, SACMEQ and PASEC) and three international agencies (UNESCO Institute of Statistics, the World Bank, the European Commission) attended the GA meeting in Hong Kong. Yet representatives from OECD, which sponsors the other main international assessment activity (PISA), did not attend.¹² Competition, regional interests and limited national capacity for assessment are major barriers to international cooperation. At the same time, considerable technical cooperation can be observed, particularly in the activities of the IEA Data Processing Center and the application of selected software for sampling, data entry and analysis. A new research institute, the IEA-ETS Research Institute, may enhance further technical cooperation.

Constraints to Global Partnerships

Both institutional and disciplinary competition affects international assessments. At present, two major institutions dominate international assessments: IEA and OECD. Both conduct studies of learning in reading, mathematics and science (and other areas), and both use many of the same technologies and experts in carrying out these assessments. One recent comparison of the IEA and OECD mathematics assessments concluded: "...there are actually a great number of similarities between them. In fact, the PISA sampling and analysis appear to have been largely lifted from the TIMSS model, refined as it has been over years of experience" (Lovelace, 2007:253). Yet, due to the high costs involved in carrying out the studies, and limited resources within countries—even high income countries—competition for these resources is inevitable. IEA has "an enviable reputation for scholarship, research and integrity over its long and distinguished career" (Lovelace, 2007: 258). Yet, as an independent nongovernmental entity it must constantly seek funding for its activities. By comparison, OECD, a membership organization of countries, has the capacity to leverage significant resources through its members' Ministries of Finance.

Additional competition comes from the World Bank, private foundations and a few research organizations that have begun supporting or developing assessments of initial competencies in reading and mathematics, targeted at the low income countries (Abadzi 2006, Crouch 2008, Pratham 2008). While few low income countries participate in IEA studies (only Ghana, Mongolia and Yemen participated in PIRLS 2006 or TIMSS 2007, and the results from Mongolia were not reported) and none participate in the OECD studies, the new assessments introduce more competition for scarce technical resources within developing countries.

Disciplinary competition over assessments is on the wane, having peaked in the 1990s with debates regarding "performance assessment" versus "multiple-choice questions" (see, for example, Shavelson, Baxter and Price 1991). The time requirements for performance assessment became overly burdensome for implementers (Black 1994), while new technologies for including and scoring "constructed response" items in paper and pencil assessments became available. Current assessments, including those conducted by IEA, typically use both constructed response and multiple-choice questions. But residuals of these debates may be found in the language

¹² The 2009 IEA GA Meeting was held in Berlin and was attended by 41 GA members, two PASA partners (PREAL and WAEC), and one international agency (UNESCO Institute for Statistics); again OECD was not represented.

contrasting IEA's "curricular" focus with OECD's "life skills" focus,¹³ contrasts that are more rhetorical than real. As Lovelace notes: If the OECD studies are supposed to be measuring "life skills", why are they "conducting assessments, in schools, at a stage before most pupils leave education and using a paper-and-pencil test administered to pupils sitting in rows in classrooms?" Other differences between IEA and OECD assessments have been noted, including the difference between a grade-level assessment and an age-level assessment, linkage of student performance with teacher characteristics and behavior, and the collaborative nature of the IEA designs (Wagemaker 2008). A recent study finds strong similarities between reading assessment in PISA and PIRLS (Grisay, Gonzales and Monseur 2009). Discussions of the differences between IEA and OECD studies can also lead to confusion in developing countries, where the technical details underlying these debates may be little understood.

Second, regional interests constrain global cooperation. PASA has provided support for four regional organizations to carry out activities related to assessment. Two of the four also carry out periodic assessments of learning achievement, following methods similar to those used by IEA (in terms of methodology, data collection procedures and analysis), one focuses on the usage of existing data for policy, while the fourth concentrates on high stakes examinations. Technical specialists from different organizations that participate in the regional assessments cooperate on an individual, personal basis, but language and geography contribute to regional hegemony.

Third, for many developing countries, national technical capacity for assessment is very limited. Staffs of assessment units are hard put to carry out all the tasks associated with rigorous large-scale assessments. One NRC mentioned this specifically: "We don't have enough staff, and the same people work on PIRLS, PISA, TIMSS." Another NRC noted that specific skills were lacking, noting that "ETS has 300 psychometricians, and we have only one." The difference in periodicity of TIMSS, PIRLS and PISA means that countries that wished to participate in both IEA and OECD studies faced significant challenges in 2006, when both PIRLS and PISA data collection took place. Assessment activities of SACMEQ and PASEC do not appear to be coordinated with the larger international assessments.

Steps to Partnership through Technical Cooperation

Despite these constraints on cooperation, aspects of global partnerships can be seen forming around the IEA and its partner institutions. Indeed, at the technical level, much cooperation among agencies exists. PASA facilitated making IEA's proprietary software available to its partners for scientific sampling of schools (W3S), for data entry and cleaning (WINDEM) and data analysis (IDB Analyzer). In addition PASA enabled IEA to provide technical assistance with these tools, which have been utilized by SACMEQ, participants in PIRLS 2006 and TIMSS 2007, and participants in training offered by the IEA-ETS Research Institute (IERI).

IEA's Data Processing Center (DPC) is a strong technical institution that provides training and services related to assessment. The DPC is currently working with six IEA and two OECD international studies and has a staff of more than 130. DPC has taken the lead in developing the software mentioned above, as well as Survey System (a system for creating on-line versions of questionnaires and administering them on line), ODA (an on-line data analyzer), and Repository Metas Search (RMS, an on-line system to download data from IEA surveys). The DPC includes a Research and Analysis Unit, which conducts research, training and analysis using IEA data. At the 2010 Annual Meeting of the Comparative and International Education Society, the IEA-DPC

¹³ In recent years OECD has softened its position with respect to "life skills" assessment in PISA.

offered two “presidential invited” workshops, one focusing on technical aspects of large scale assessments and the other on the use of TIMSS and IRLS for secondary analysis.

IEA also has taken steps in creating an international assessment institution through the establishment of a virtual research institute, the IEA-ETS Research Institute (IERI). According to the IERI website:

“The IERI is a collaborative effort between the Research & Development Division at ETS and the IEA Data Processing and Research Center that focuses on improving the science of large-scale assessments. The IERI undertakes activities around three broad areas of work that include research studies related to the development and implementation of large-scale assessments (research area); professional development and training (training area); and dissemination of research findings and information gathered through large-scale assessments (dissemination area)...Research projects are hosted in the funding institutions, but are facilitated by Web-based collaboration as well as by the shared joint expertise of researchers involved in work on large-scale assessments. The aim of this virtual research area is to contribute to the science of large-scale assessments so that the best available information is provided to policy makers and researchers from around the world.”

The IERI has offered two types training academies over the past two years, one for the analysis of a specific data set and a second type addressing technical topics. In May 2008, an academy was held to analyze the PIRLS 2006 and PISA 2006 data sets; in May 2009 an academy was devoted to training participants in analyzing the TIMSS 2007 data; and in May 2010 an academy will be held on analyzing TIMSS 2007, PIRLS 2006 and PISA 2006. Academies devoted to specific topics were held in November 2007 for HLM analysis and in November 2008, January 2009 and November 2009 for assessment design, IRT and plausible values. Participants in these academies included technical specialists and researchers from 13 developing countries.¹⁴ The IERI has also published two volumes of the IERI Monograph Series on *Issues and Methodologies in Large Scale Assessments* (von Davier and Hastedt, Eds, 2008, 2009). Authors in these volumes include many individuals associated with IEA studies. In addition, IEA has initiated conversations with several universities, with respect to post-doctoral fellowships and internships.

Conclusion

Progress toward achieving a global partnership for assessment is gradually being made, despite significant constraints. No concrete international institution for assessments has been formed, and no sustained funding for assessment has been achieved. The existing international and regional partnerships for conducting specific studies are consortia rather than concrete institutions and are dependent upon donor grants and participant fees. A concrete institution for assessment would have qualities identified as essential for other applied research institutions; it would be stable, staffed by professionals who have received high-level doctoral, postgraduate training, provide master’s level training opportunities for junior staff and such “extension workers” as test administrators, and enable its professional staff to communicate with their peers internationally (Larach and Lockheed 1992). Such institutions exist at the national but not the global level; there remains a need for an International Institute for Assessment.

¹⁴ Ghana, Hungary, Iran, Jamaica, Kazakhstan, Lithuania, Poland, Romania, Russian Federation, Slovak Republic, South Africa, Ukraine, and Yemen.

Such an institute could be built on the framework already in place with IEA, including the IEA Data Processing Center (DPC), the IEA-ETS Research Institute, PASA partner institutions, other regional partnerships such as the Organization of Ibero-american States (OIS), and universities. Such a network could well constitute a global partnership for assessment and could be superior to a single entity with narrower mandate (such as simply training or data collection). At the same time, such an institute would require sustained, long-term funding for it to be viable.

Chapter 6: PASA and Education Policy

Results from the two IEA studies (PIRLS 2006 and TIMSS 2007 and their predecessors) that were supported through PASA and the activities of the four PASA regional partners (PASEQ, PREAL, SACMEQ, and WEAC) have the potential to affect educational policy in the participating countries, particularly in terms of leveraging national interest in learning outcomes, refining curricula and improving teacher training programs (Wolff 2007, Kamens and McNeely 2009). Internationally, the results of earlier PIRLS and TIMSS have been reported in all recent UNESCO Global Monitoring Reports and used to support a variety of policy recommendations in these documents. Background papers for the current GMR 2010 include reference to PIRLS 2006 and TIMSS 2007 (Altinok 2009).

All major international and regional assessments have generated a literature related to education quality; a recent search suggests that the IEA studies (TIMSS and PIRLS) have generated the most scholarly reports (table 7). In addition, PIRLS 2006 and TIMSS 2007 were visible at the 2010 Annual Meeting of the Comparative and International Education Society, where the distinguished Eggertsen Lecture was delivered by the head of the PIRLS and TIMSS International Study Center at Boston College and numerous panels features secondary analyses of these large-scale surveys.

Table 7: IEA Studies (TIMSS and PIRLS) are the Most Referenced International or Regional Assessments

Assessment	Google Scholar Pages 2009	Google Scholar Pages 2010
IEA (TIMSS and PIRLS)	21,480	24,160
OECD (PISA)	16,300	22,000
LLECE	796	1,560
PASEC	664	800
SACMEQ	521	729

Source: Google Scholar, February 12, 2009 and February 25, 2010

IEA Assessments and Education Policy

The impact on education policy of the two major studies carried out by IEA under the PASA umbrella – PIRLS 2006 and TIMSS 2007 – was difficult to assess at the outset of this evaluation, a concern raised in the previous evaluations that focused on TIMSS 1999 (Elley 2002) and PIRLS 2001 and TIMSS 2003 (Gilmore 2005). Since the results of PIRLS 2006 became available only in 2007 and the results of TIMSS 2007 in December, 2008, few IEA General Assembly (GA) members and TIMSS 2007 national coordinators from developing countries mentioned any specific policy impacts of these assessments during the interviews in October or December, 2007. When the closing date of the grant period was extended to December 2009, however, an opportunity to interview TIMSS 2011 national coordinators arose; these interviews suggested areas of policy impact. Policy impact can also be inferred from the rise in visibility of the IEA large scale assessments in the academic literature, professional conferences and international policy documents.

Interviews at the end of 2007 underscored the positive effect of prior assessments on one aspect of education policy: curriculum reform. For example, representatives from developing countries noted the following:

- “TIMSS 1995 had an impact on science and math education in the country; it provoked a reform based on the TIMSS framework leading to a big improvement from 1995-1999.”

- “Education experts want to improve education. PIRLS 2001 led to change in textbooks; a new textbook was written. PIRLS 2006 shows improvement over 2001, perhaps due to the change in the textbook.”
- “The results of TIMSS 2003 were presented to the Cabinet and funds were received for improving science education.”
- “The report [TIMSS 2003] was submitted to government bodies, universities, Board of Education. The result was that the basic education curriculum was updated, including in-service training of teachers on new curriculum.”
- “Obvious curriculum shortcomings are being addressed in ongoing curriculum revisions.”

A few developing countries participated in both TIMSS 2003 and TIMSS 2007, and a comparison of their scores over time could indicate whether the earlier assessment had an impact on curriculum revisions leading to higher achievement. However, at both grades approximately the same number of countries showed improvement as showed decline. For example (statistically significant changes are given in parentheses) on the 4th grade math assessment, four countries improved: Armenia (+ 44 points), Iran (+ 13 points), Latvia and Russia; and four countries declined: Hungary (- 19 points), Lithuania, Morocco and Tunisia (- 13 points). On the 8th grade math assessment, eight countries improved: Armenia (+ 21 points), Ghana (+ 34 points), Jordan, Lebanon (+16 points), Lithuania, Russia, Serbia (+9 points) and Tunisia (+ 10 points); and eight countries declined: Botswana, Bulgaria, Egypt (- 16 points), Hungary (- 12 points), Indonesia, Iran, Malaysia (-34 points), Romania (-14 points) and Serbia (- 23 points). Several interviews suggested that curriculum change has a longer implementation period than the time between two international assessments, and some countries report planning to gauge the effects of curriculum reform on the TIMSS 2011 assessments. Reports from GA members in October 2008 noted that national improvements on PIRLS 2006 as compared with PIRLS 2001 demonstrated the impact of earlier curriculum reforms that addressed shortcomings in reading literacy.

Interviews in 2009 reported that some discussions of education policy as related to TIMSS 2007 had occurred in their countries, although several of those interviewed reported that TIMSS 2007 results had not yet been widely disseminated. This was particularly true for low- and lower middle-income countries and for those participating for the first time:

- “When reports came out, elections were going on, so not so much attention to results. National report is ‘in press.’”
- “National report has not yet been discussed.”
- “Results have not been unveiled.”
- “No newspaper report on results.”
- “No discussion; when we had results we were very busy with teacher exams and provided nothing for the media.”

When discussed, policy reforms mentioned by the 2011 national study coordinators focused largely on teacher training, curriculum reform and assessment reform:

- “State Secretary of Education wants to use results for policy, and will establish a professional group for analysis of each subject. Secondary analyses to find out why [country name] did not score ‘so good.’”
- “Lots of discussion of teaching and learning at the National Center of Education about reforms of assessment toward continuous assessment and combining learning and testing.”
- “In 2003, people started to have interest, to compare [country name] with other countries. Curriculum people went around to see primary and lower secondary curriculum frameworks in other countries.”
- “The Ministry asked for public discussion of results and approved a workshop on how to use the methodology of international assessments in national assessments.”
- “The Ministry paid more attention to the results, which was the first time in 30 years, and the Deputy Minister is chair of a working group.”
- “Curriculum updated at all levels, textbooks upgraded, and teachers upgraded with in-service training.”

The interviews also indicated that results can have budgetary effects on assessment institutions, both positive and negative. For example:

- “For PIRLS, the director of the institute said they could not afford a press release about results, for fear of losing funding from the Ministry of Education. In the past, financial support was cut as a result of poor results.”
- Has led to increasing financing from government

Other interviewees noted that trend data, in particular, were important and that assessment data were used to support justification of projects for World Bank support:

- “TIMSS 2003 was the first assessment; the result was average and there was not response from the Ministry of Education. Having trends from TIMSS 2007 will be important.”
- “PIRLS data were used in preparation of projects for World Bank funding.”

An interviewee from the World Bank noted that the World Bank “encouraged countries to participate in PIRLS and TIMSS because they are curriculum-based, with direct policy implications for curriculum reform” and that the results from the analyses of PIRLS and TIMSS were used in developing World Bank projects. The MENA seminars were particularly targeted at developing skills for analyzing and writing policy relevant documents based on the IEA studies. During the seminars, participants carried out secondary analyses of the earlier PIRLS and TIMSS data, and a World Bank interviewee noted that the results of such analyses stimulated policy change vis-à-vis the curriculum in several countries. Moreover, the results from earlier IEA studies have been used by multilateral donors in their own policy reports. The MENA Region’s recent “flagship” report, *The Road Not Travelled: Education Reform in the Middle-East and*

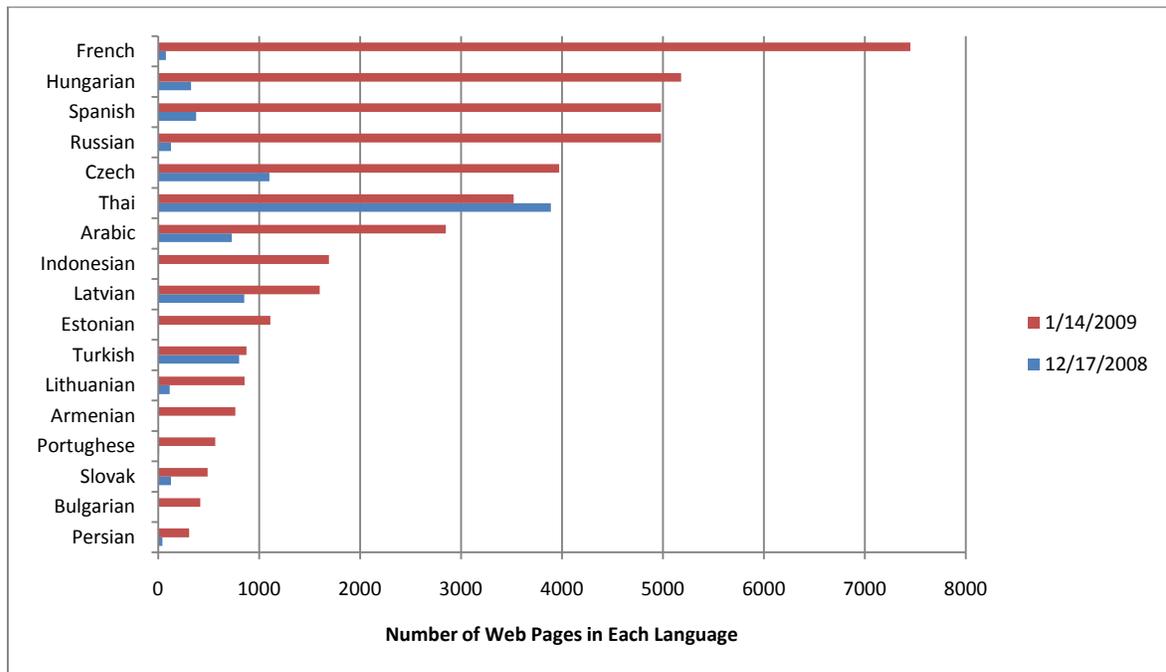
Africa (World Bank 2008) has a chapter that uses TIMSS 2003 data at both country and regional levels. The UNDP produced a report for the MENA region based on PIRLS.

Presentations by IEA members at the 2008 GA meeting identified some policy changes that derived from analyses of PIRLS 2006. Specifically, France, Hong Kong, Iceland and Russia reported how PIRLS 2006 results had stimulated discussion of curriculum and textbook reform. Recent newspaper articles have also indicated that results from TIMSS 2007 have generated discussion within countries; for example, in Malaysia the results prompted a public discussion about the policy for mathematics and science to be taught in English rather than in Bahasa Malay.

Another indicator of the extent to which TIMSS 2007 may affect policy is provided by its visibility as indicated by the number of web pages, in different languages, that reference this study. We compare this number extracted from the web around the time of the TIMSS 2007 release in December 2008 with the number of pages extracted a month later, after the results of the study had been more widely disseminated. Overall, visibility of TIMSS 2007 increased dramatically over this month, with the number of web pages in all languages leaping from 22,900 pages in December 2008 to 194,000 pages in January 2009 – an eight-fold increase. In English alone, the number of pages jumped from 8,380 to 63,100.

TIMSS 2007 visibility in nearly all of the other languages used by the developing countries that participated in TIMSS increased even more dramatically; the figure below compares the number of web pages referencing TIMSS 2007 in December, 2008 with the number referencing TIMSS 2007 a month later. While visibility on the web does not ensure that TIMSS 2007 results will be used for policy, it does suggest that information on the learning performance of students is reaching an audience in the developing countries participating in the study. Such information is a requisite for evidence-based policy recommendations.

Figure 5: Web pages referencing TIMSS 2007 increased from December 2008 to January 2009.



Source: Author's extraction of pages referencing TIMSS 2007 in specific languages, using Internet Explorer and Google on December 17, 2008 and January 14, 2009.

PASA Partners and Education Policy

Publications and anecdotal reports regarding the impact on educational policy from the activities of the PASA partners suggest that all partners have played some part in curricular analysis and reform in the countries in which they have been working. In addition, two partners (PREAL and SACMEQ) appear to have contributed significantly to policy dialogue and reform in their respective regions, while two other partners (PASEC and WEAC) have focused more narrowly on curricular analysis, testing methodologies and training relative to these topics.

PREAL's activities have included studies, meetings and conferences, publication and training designed to increase the visibility and use of assessments in education policy in Latin America. Completed publications include a manual on national education assessments, a manual on education standards, an overview of the development and prospects of assessment in Latin America, a paper on the costs of educational assessments, a paper on the technical challenges of large-scale evaluations, a paper on "The Educational Assessments that Latin America Needs", five country case studies on educational standards in Latin America, policy briefs on such topics as the costs of educational assessments and the results from PIRLS. Training on curricular analysis and test design and on the use of tests took place, in partnership with WBI. Such publications and training have the potential of affecting education policy, as well as are concrete activities in support of a global partnership in assessment.

SACMEQ's assessments are designed to address specific policy questions identified by the ministers of education of participating countries. While PASA funds have supported the preparation and data collection for SACMEQ III, results from SACMEQ II have informed education policy. For example, Murimba notes that "the greatest impact SACMEQ has had on ministries of education is on policy-related, system development processes" (2005:98). The paper provides examples of SACMEQ's influence on ministry understanding of the discrepancies between official norms for schools inputs and the actuality of these inputs in Zimbabwe, Kenya, Malawi, and the Seychelles. It also revealed discrepancies between actual practice and "policy rhetoric" regarding school fees and repetition in all participating countries, leading to policy reform in some cases. Data from SACMEQ have informed such policy documents as national education plans, UNESCO EFA Global Monitoring Reports, and the World Bank's World Development Reports; have been used in the preparation of education projects by international donors; and have been analyzed by education researchers pursuing master's and doctoral degrees.

Conclusion

The impact on education policy of the two major studies carried out by IEA under the PASA umbrella – PIRLS 2006 and TIMSS 2007 – is difficult to assess, since the results of PIRLS 2006 became available only in late 2007 and the results of TIMSS 2007 in December, 2008. Some interviews, however, underscored the positive effects of earlier assessment results on three aspects of education policy: reforms in curriculum, assessment and teacher education.

The public visibility and academic use of these international studies has sharply increased over time. Web page references to TIMSS, for example, jumped sharply immediately after the release of the international report in December 2008. The academic literature referencing these studies, as well as their visibility at professional conferences and international reports, has also grown, which may, in the long run, affect education policy in participating countries.

The impact on policy of the activities of two PASA partners, PREAL and SACMEQ, is also difficult to assess, but both have contributed to dialogue on assessment within their respective

regions. In the case of SACMEQ, the assessment instruments have been designed to address the educational policy concerns of participating countries, as discussed at regular meeting of Ministers of Education in the region. In the case of PREAL, substantial progress has been made in raising awareness regarding the importance of assessment for improving the quality and effectiveness of education in Latin America (Ravela and others 2008) and in training in the use of assessment evidence to inform policy.

Chapter 7: Conclusions and Recommendations

The conclusions of the previous independent evaluations of IEA and earlier TIMSS assessments, specifically, noted that: (a) the participation of low and middle-income countries in international assessments would not have been possible without direct financial support for that purpose, (b) such participation helped develop specific technical and administrative skills related to large-scale assessments, and (c) the results from the assessments generated interest and, in some cases, action related to curricular reform (Elley 2002, Gilmore 2004, Aggarwala 2004).

The recommendations of the previous independent evaluations were remarkably consistent, emphasizing continued World Bank support to IEA, continued emphasis of training for NRCs, greater attention to inclusion of low and middle-income countries, particularly large ones such as China, India, Nigeria, Brazil and Argentina, and targeted, supplemental training for “novice” countries. These messages again emerge from the present evaluation.

Conclusions

The PASA program was designed to address the recommendations of earlier evaluations, and IEA generally implemented PASA in accordance with its plan, despite significant reductions in the size and duration of the DGF award relative to IEA's initial proposal. PASA support enabled 19 developing countries to participate in and complete PIRLS 2006 and/or TIMSS 2007, which increased the number of such countries reporting measures of learning achievement for primary and lower secondary school age children; it also exposed another eight countries to methodologies for assessment, even though they did not complete the exercise. IEA's website is being used, and usage increased substantially, from 2006 to 2008. PASA's program in training for the MENA region was completed as planned, with participation from 14 countries and stable institutional participation. Expectations for IEA's partnership with WBI in training related to national assessments did not materialize, although experts from some institutions associated with IEA were involved in the training; substantial training was provided through IEA for developing countries. Uncertainty regarding funding and the weakness of the US\$ against the Euro created a financial hardship for IEA in its efforts to include developing countries in the international assessments.

Direct training and participation in various IEA activities has contributed to improved capacity for assessments, as indicated by participant self-reports, regularity of participation in international assessments and regularity of national assessments. Some alignment between those participating in the IEA studies and those responsible for national assessments was observed. The lack of information on the actual quality of national assessments (in contrast to the availability of substantial technical details on the quality of international assessments) means that it is not possible to determine the impact of PASA on the quality of national assessments.

An innovative feature of PASA was its focus on building an international community of expertise in assessment, though the support of and sustained contact with its regional partner organizations. PASA's impact on the four regional assessment and policy analysis partners has been positive although varied, depending on the objective of each organization. For **SACMEQ**, PASA provided financial continuity at a time of disruption, which enabled SACMEQ III to proceed, approximately on schedule. For **PASEC**, PASA supported essential activities that served as a foundation for regional assessments and broke new ground; it also enabled PASEC to establish and maintain contact with experts from the other PASA partners and built its own capacity for assessment. For **WAEC**, PASA support enabled the training to improve continuous assessment in schools and strengthen technical skills for staff working on a secondary school leaving

examination. For **PREAL**, the PASA support enabled the production of documents related to the quality and cost of national assessments and to deliver training programs in assessment for experts and government officials in Latin America; it also enhanced the dialogue about assessment in the region through its publications and training.

Progress toward achieving a global partnership for assessment is gradually being made, despite significant constraints. Regional groups, such as the PASA partners, have a broader awareness of assessment activities in other regions and have benefitted in a variety of technical ways from participating in PASA. These contacts are fragile, however, and dependent upon external financial and professional support. No concrete international institution for assessments has been formed. The existing international and regional partnerships for conducting specific studies are consortia rather than concrete institutions. A concrete institution for assessment would have qualities identified as essential for other applied research institutions; it would be stable, staffed by professionals who have received high-level doctoral, postgraduate training, provide master's level training opportunities for junior staff and such "extension workers" as test administrators, and enable its professional staff to communicate with their peers internationally (Larach and Lockheed 1992). Such institutions exist at the national but not the global level, unlike entities with related mandates such as UNESCO's International Institute for Educational Planning (IIEP) and Institute for Statistics (UIS).

There remains a need for an International Institute for Assessment. Such an institute could be built on the framework already in place with IEA, including the IEA Data Processing Center (DPC), the IEA-ETS Research Institute, PASA partner institutions, other regional partnerships such as the Organization of Iberoamerican States (OIS) with whom IEA has recently signed an agreement, and universities. Such a network could well constitute a global partnership for assessment, could be superior to a single entity with a narrower mandate (such as simply training or simply data collection) and could constitute a "global public good."

Finally, the impact on education policy of the two major studies carried out by IEA under the PASA umbrella – PIRLS 2006 and TIMSS 2007 – and of the activities of the PASA partners is difficult to assess, given the time frames involved. The results of PIRLS 2006 became available only in 2007 and the results of TIMSS 2007 in December, 2008; SACMEQ III data has only just been collected and data entry is in progress. Some interviews and reports, however, underscored the positive effect of earlier assessment results on education policy; PIRLS and TIMSS stimulated curriculum reform in many countries and SACMEQ provided evidence regarding discrepancies between "policy rhetoric" and actual practice. The activities of two PASA partners, PREAL and SACMEQ have contributed to dialogue on assessment within their respective regions and, in the case of SACMEQ, the assessment instruments have been designed to address the educational policy concerns of participating countries. Moreover, the visibility of these studies has sharply increased over time, as indicated by web page references, and the academic literature has benefitted from the studies, which may, in the long run, affect education policy.

Recommendations

Consistent with previous evaluations, this evaluation recommends:

- *Continued World Bank support to IEA* to offset the costs to low- and middle-income countries who participate in upcoming international assessments, particularly PIRLS 2011 and TIMSS 2011. The support should extend for the duration of the assessment cycle.

- *Continued training by IEA* for participants in such studies, including supplemental training (such as the MENA training) for participants in low- and middle-income countries.
- *Continued encouragement of larger developing countries to participate in IEA studies*, and, in some cases, to bring their expertise to the table.

In addition, this evaluation recommends:

- *Support for an international institution* that would both conduct assessments and train the next generation of researchers, to build capacity for large-scale national and international assessments and provide a “global public good.” IEA’s model of “learning by doing” has proven to be effective, whereas single-purpose training or research institutions do not provide the opportunity to practice, in a real-life situation, what is being taught. IEA and its DPC could serve as a basis for such an international institution.

Finally, this evaluation notes that very few low-income countries participate in international assessments and those that do report considerable difficulties in meeting the standards of implementation.¹⁵ Thus, this evaluation recommends:

- *IEA concentrate on building capacity among lower-middle income countries* rather than intensifying efforts for low-income countries.
- *Regional assessment institutions continue to work with low-income countries* to build their assessment capacity through training and participation in regional assessments and studies.

¹⁵ No low-income countries participated in PIRLS 2006 and only Ghana, Mongolia and Yemen participated in TIMSS 2007 (Mongolia’s results were not published). Even some lower-middle income countries have difficulty in meeting the implementation standards of IEA.

References

- Abadzi, H. 2006. *Efficient Learning for the Poor*. Washington, DC: World Bank
- Aggarwala N. 2004. *Quality Assessment of Primary and Middle Education in Mathematics and Science (TIMSS)*. Takoma Park, MD: Eaton-Batson International
- Altinok, N. 2009. "An Empirical Approach to Marginalization in Education based on the TIMSS 2007 Study". Paper commissioned for the *EFA Global Monitoring Report 2010, Reaching the Marginalized*. Paris: UNESCO.
- Black, P. 1994. "Performance Assessment and Accountability: The Experience in England and Wales." *Educational Evaluation and Policy Analysis* 16 (2): 191-203.
- Crouch, L. 2008. Presentation at Early Grade Reading (EGRA) workshop, March 12-14, Washington, DC.
- Elley, W. 2002. *Evaluating the Impact of TIMSS-R (1999) in Low- and Middle-Income Countries: An independent report on the value of World Bank support for an international survey of achievement in mathematics and science*. New Zealand.
- Ferrer, G. 2006. *Educational Assessment Systems in Latin America: Current Practice and Future Challenges*. Washington, DC: PREAL
- Galiani, S. and J. Corrales. 2006. "Academic Evaluation Offices in Latin America: An index of institutional capacity"
- 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.
- Grisay, A., E. Gonzalez and C. Monseur. 2009. "Equivalence of item difficulties across national versions of the PIRLS and PISA reading assessments." *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments*. " IERI.
- Kamens, D. and C. McNeely. 2009. "Globalization and the Growth of International Educational Testing and National Assessment." *Coomparative Education Review* 54 (1): 5-25.
- Larach, L. and M. Lockheed. 1992. *World Bank Lending for Educational Testing: A General Operational Review*. Washington, DC: The World Bank.
- Lietz, P., H. Wagemaker, O. Neuschmidt & J. Hencke. 2008. *Issues in the Middle East and North Africa Region: Outcomes of the IEA Arab Region Training Seminar Series 2006/2007*.
- Lockheed, M. 2008. *Measuring Progress with Tests of Learning: Pros and Cons for Progress-based Aid in Education*. CGD Working Paper 147. Washington, DC: Center for Global Development
- Loveless, T., Ed. 2007. *Lessons Learned: What International Assessments Tell Us About Math Achievement*. Washington, DC: Brookings
- Marimba, S. 2005. "The impact of the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ)." *Prospects* 35 (1): 91-108.

- Pratham. 2008. *Annual Status of Education Report (Rural) 2007 Provisional*. Delhi: Pratham.
- Ravela, P., P. Arregui, G. Valverde, R. Wolfe, G. Ferrer, F. Martínez Rizo, M. Aylwin, and L. Wolff. 2008. *The Educational Assessments that Latin America Needs*. Washington, DC: PREAL.
- Ross, K. and I. Genevois, Eds. 2006. *Cross-national Studies of the Quality of Education: Planning their design and managing their impact*. Paris: IIEP
- Shavelson, R., G. Baxter, J. Pine. 1991. "Performance Assessment in Science." *Applied Measurement in Education* 4(4): 347-362.
- UNESCO. 2007. *EFA Global Monitoring Report: Education for All by 2015—Will We Make It?* Paris: UNESCO.
- von Davier, M. and D. Hastedt. 2009. *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments, volume 1*.
- Wagemaker, H. 2008. "Choices and trade-offs: reply to McGaw." *Assessment in Education: Principles, Policy and Practice* 15 (3): 267-278.
- Wolff, L. 2007. "The Costs of Student Assessments in Latin America". PREAL Working Paper No. 38. Washington, DC: PREAL
- World Bank. 2004. *Books, Buildings and Learning Outcomes: An impact evaluation of World Bank support to basic education in Ghana*. Washington, DC: World Bank.
- World Bank. 2006. *From Schooling Access to Learning Outcomes: An Unfinished Agenda*. Washington DC: World Bank.
- World Bank. 2008. *The Road Not Travelled: Education Reform in the Middle-East and Africa*. Washington, DC: World Bank

Annex A: PASA Objectives and Program Components

Component 1. Strengthen Capacity for International Assessment

The program will support IEA to do the following in the context of developing and conducting international assessment.

1. Capacity Building and Institutional Development

- Train participant country staff members in educational assessment – sampling, data management and analysis; reporting, training and assistance with national reporting and analysis.
- Train policy and management teams on evidence-based policy making, on the management and organizational requirements for assessment; and on institutional requirements for sustainable assessment systems.
- Conduct seminars with a focus on integrating policy issues with the use of data from international assessments.
- Liaise extensively with Bank operational staff in carrying out these activities and exploring operational linkages.

2. Data Production and Analysis

- Provide valid and reliable data on mathematics, science and reading literacy for over 15 countries as per the schedule for Progress in International Reading Literacy Study (PIRLS), and for Trends in Mathematics and Science Study (TIMSS). This will include international comparable achievement data at grades 4 and 8, international achievement benchmarks, trend data in educational achievement, and international and national data sets for secondary analyses.

3. Technical Support at the international level

- Share tools, methodologies, good practices and approaches among participant countries.
- Provide on line data analysis tools.

4. Region-Specific Support

- As part of a first effort in supporting capacity building for regions involved in the international assessment, it will hold regional workshops for MNA to bring the MNA countries together to explore collaboration and coordination, standard setting within the region, the sharing of information, cross-regional comparison, and the development of a community of experts and knowledge within the region. It will also support the development of national regional reports for policy development.
- For participants of the WBI regional program on capacity building for national assessment in Africa and South Asia - it will provide follow-up training for groups of countries, enhanced data quality using IEA software programs, provide analytical support for the comparison of results, train national technical experts for training delivery, and support conference of participant countries on national assessment and issues involved in institutional development of educational assessment.
- It will provide a forum for the coordination and harmonization of regional assessment programs such as those supported by this DGF under Component 2.

Component 2: Strengthen Capacity of Regional Institutions for Educational Assessment and Evaluation

Four regional assessment institutions from Africa and LAC are participating in the program and contributed to the overall PASA design.

5. Regional Capacity Building in Africa

In Africa, the program supports the quality and sustainability of the regional assessment programs in Francophone and Anglophone Africa. The programs are developed by non-profit inter-governmental organizations dedicated to educational assessment, research and training. World Bank support in the past to these groups has been ad hoc, and not based on a coherent strategic framework for developing educational assessment and its use for policy and program improvement. PASA has allowed the Bank's regional staff to be engaged in a more systematic basis with these regional programs.

- The Programme d'Analyse des Systems Educatifs de la Confemen (PASEC), www.confemen.org is a program of the Conference of Ministers of Education in Francophone Countries (CONFEMEN). It was launched in 1991 with the aim of providing decision-makers in developing member countries with objective information on their basic education systems to facilitate the development of efficient education policies. PASEC conducts quality assessment in primary education, trains national teams for carrying out assessment, and diffuses results as well as assessment and analytical tools and models for use at the national, regional and global levels. In Sub-Saharan Africa and the Indian Ocean, the CONFEMEN counts 25 member countries, 15 of which have already benefited from an assessment. PASA support is used to: (i) upgrade its technical capacity for assessment to comply with the international standards prescribed for comparison of results between countries; (ii) initiate partnerships with other regional assessment programs to ensure better comparability of results; (iii) begin to carry out in French the WBI training program on national assessment in CONFEMEN African member countries. Support for the latter is for the development of training material (adaptation into French and production), organization of training sessions, and technical assistance (co-ordination and training).
- The Southern Africa Consortium for the Measurement of Educational Quality (SACMEQ) <http://www.sacmeq.org> is a network of 15 Ministries of Education located in Southern and Eastern Africa. Its main mission is to expand opportunities for educational planners and researchers to gain technical skills required to monitor and evaluate the quality of their basic education systems, and to generate information that can be used by decision-makers to plan and improve the quality of education. PASA was intended to strengthen the SACMEQ Coordinating Center (SCC) to: (i) provide adequate technical and managerial support to SACMEQ Ministries; (ii) provide technical training to its members countries; (iii) expand its research and training programs to include a much wider range of educational planners and researchers, (iv) enhance effective use of the SACMEQ data base for global use by the development of data systems and archive, (v) develop websites and provide training to academic staff, researchers, and policy makers on the use of the SACMEQ data archives; (vi) share models and methodologies for adoption or adaptation among member countries.

- For the West African Examinations Council (WAEC) with interest in expanding its role to include national assessment, the focus is on a comprehensive analysis of its capacity for educational assessment covering both examinational and national assessment of Anglo-phone West African countries.
- For all these Africa Region institutions, the IEA will convene a plenary meeting to identify common issues related to standard setting, and to develop plans for collaboration in capacity building and harmonization.

6. Regional Capacity Building in Latin America and the Caribbean (LAC)

For the LAC region, the program will support the Partnership for Educational Revitalization in the Americas (PREAL) to enhance the effective use of education assessment data collected from national assessment initiatives. PREAL is a region-wide network of institutions seeking to improve education policy. It was established in 1995 by the Inter-American Dialogue (in Washington) and CINDA (in Santiago, Chile) as a multi-year initiative aimed at building a broad and active constituency for education reform in many countries. Since then it has become the leading non-governmental voice on education in Latin America and a strong advocate for involving leaders from civil society in the work of education reform. A majority of PREAL's activities are carried out in collaboration with expert public policy centers throughout the region working to promote education reform.

Using assessment information to guide policies, programs and investments decisions is noted as one of the major weaknesses in the development of assessment systems. This component with the LAC region provides an opportunity to do just this type of work. PASA provided PREAL support to work with over 15 LA countries that have completed national assessments under Bank financed projects to carry out the following activities:

- Hold conferences and seminars to expand the region-wide network of specialists and professionals for a more active involvement in educational assessment and the use of assessment data to conduct studies of policy relevance to LAC;
- Prepare reports and publications on good practices in assessment;
- Expand the on-line clearinghouse on assessment to include translated information from all over the world;
- Adapt and translate existing training materials to meet regional demands and contexts.

Annex B: Evaluation Questions and Sources of Data

Evaluation Question	Details	Data Sources
Has IEA implemented PASA as planned?	Developing countries participate in the 2007 TIMSS and the 2006 PIRLS (PASA supports training/meeting costs, data management, translation verification, data use and policy analysis seminars)	IEA documents on country participation in TIMSS 2007 and PIRLS 2006 (including training and meetings). TIMSS 2007 progress reports . IEA Annual Reports , 2006 and 2007. IEA Audits for 2006 and 2007. Interviews with Executive Director of IEA and IEA officers Post-training evaluations completed by participants in IEA/PASA training and lists of attendees at training sessions, identified by organization and country
	IEA website expansion	Usage information from IEA website (number unique visitors, visits, hits; country of visitors; downloads of test items, materials, tools).
	Regional seminar for assessment institutes in countries of the Middle East and North Africa Region	IEA and WB documents on MENA training. Interviews with WB education sector staff, particularly from MENA.
	World Bank Institute program on assessment	IEA and WBI documents on WBI program. World Bank records of knowledge and learning products related to assessment (FY06-08)
Has PASA improved the capacity of developing countries in the area of assessment of learning?	More lower- and middle-income countries participate in TIMSS 2007 and PIRLS 2006 than in TIMSS 2003 and PIRLS 2003; more lower and middle income countries participate in other international assessment (e.g. PISA); developing countries report greater capacity	Review of national reports for TIMSS/PIRLS, other international assessments. Interviews with IEA representatives from developing countries (10 GA representatives, 26 TIMSS 2007 study coordinators from 28 developing countries).*
	National body for assessment exists in more LICs and MICs	Review of national body for educational assessment (staffing, authorizing framework) Budget support for national assessment
Have PASA partners (regional assessment bodies) undertaken activities or developed products designed to enhance the capacity of national or regional assessment institutions?	PASEC (training materials in French, training sessions for CONFEMEN countries)	Interview with representative of PASEC Reviews of on-line information Annual progress reports to IEA from PASEC
	SACMEQ (training related to SACMEQ III)	Interviews with representative of SACMEQ Reviews of on-line information Annual progress reports to IEA from SACMEQ
	WAEC (training of test developers, item writers and examiners)	Interviews with representative of WAEC Reviews of on-line information Annual progress reports to IEA from WAEC
	PREAL (conference with 15 countries receiving WB support for assessments; reports and publications; on-line Spanish clearing house on assessment; training materials in Spanish)	Interviews with representative of PREAL Reviews of on-line information Annual progress reports to IEA from PREAL
Has IEA taken steps to establish a global partnership for assessment?	Meetings of potential partners have taken place	Interviews with representatives of SACMEQ, PASEC, WAEC and PREAL GA meeting agendas and minutes
Has PASA affected educational policy in developing countries	Public discussion of results in national news media Use of IEA study data in preparation of donor-supported education projects Use of IEA study data in policy research	GA meeting reports Newspaper articles Web pages Review of World Bank education projects Review of education policy literature Interviews with TIMSS 2011 study coordinators from 19 developing countries

*GA representatives interviewed (Hong Kong October 7-11, 2007): Latvia, Lithuania, Kazakhstan, *Russia*, Botswana, Indonesia, *Philippines*, *Turkey*, Malaysia, Armenia. Developing countries in italics are self-supported.

Annex C: Low- and Middle-Income Countries Participating in PIRLS 2006 and TIMSS 2007 and receiving fee support from PASA or UNDP*

Country	Results Reported in International Report			PASA / UNDP Support
	PIRLS 2006	TIMSS 2007 Grade 4	TIMSS 2007 Grade 8	
Support for low- and middle-income countries completing assessment				\$4,533,000
Algeria		√	√	UNDP
Armenia		√	√	\$240,000
Bosnia and Herzegovina			√	\$160,000
Botswana			√	\$140,000
Bulgaria	√		√	\$280,000
Colombia		√	√	
Czech Republic		√	√	
Egypt			√	UNDP
El Salvador		√	√	
Georgia	√	√	√	\$360,000
Ghana			√	\$160,000
Hungary	√	√	√	\$360,000
Indonesia	√		√	\$280,000
Iran, Islamic Rep. of	√	√	√	
Jordan			√	UNDP
Kazakhstan		√		\$160,000
Latvia	√	√	√	\$280,000
Lebanon			√	UNDP
Lithuania	√	√	√	\$313,000
Macedonia, Rep. Of	√			\$120,000
Malaysia			√	\$160,000
Moldova	√			\$360,000
Mongolia**		√**	√**	\$240,000
Morocco	√	√	√	\$360,000
Oman			√	
Palestinian Nat'l Auth.			√	UNDP
Poland	√			\$120,000
Romania	√		√	\$280,000
Russian Federation	√	√	√	
Serbia			√	\$160,000
Slovak Republic	√	√		
South Africa	√			
Syrian Arab Republic			√	UNDP
Thailand			√	
Trinidad and Tobago	√			
Tunisia		√	√	
Turkey			√	
Ukraine		√	√	
Yemen		√		UNDP
Support for low and middle-income countries not completing assessment				\$560,000
Albania	√			\$70,000
Belarus	√			\$30,000
China	√			\$60,000
Honduras		√	√	\$120,000
Nicaragua	√			\$60,000
South Africa			√	\$80,000
Uzbekistan			√	\$80,000
Zimbabwe	√			\$60,000
Support to high income country				\$147,000
Slovenia	√	√	√	
Grand total PASA support				\$5,240,000

*Excludes support for international travel/meetings, training sessions at regional seminars, incidental national costs.

** Completed assessment but results not published in international reports.

Annex D. Framework for Measuring National Capacity for Large-Scale Assessment

Capacity Dimension	Lowest Capacity	Modest Capacity	Highest Capacity
A. Technical Capacity			
Choosing curricular areas and domains for assessment	No national agency * selects curricular areas for assessment	Curricular areas for assessment are selected without reference to a broader framework for assessment	A national agency adopts an assessment framework and analyzes the curriculum and curriculum standards as basis for assessment
Item writing (forced choice items)	No national agency writes forced choice achievement items	A national agency organizes item writing for national assessments or national examinations, writes items, identifies unique correct answers for forced response items	A national agency writes and field tests forced choice items, and builds an item bank of pilot- tested items having unique correct responses
Item writing (constructed response items)	No national agency writes constructed response achievement items	A national agency organizes item writing for national assessments or national examinations, writes items and prepares scoring rubrics (guides) for constructed (open-ended) questions, including examples	A national agency writes and field tests constructed response items, and builds an item bank of pilot-tested items
Test development	No national agency assembles items into achievement tests	A national agency develops an assessment or examination framework and constructs tests based on the framework, without field testing of items	A national agency develops an assessment or examination framework, constructs tests based on the framework and empirical field test results (using modern psychometric techniques including measures of bias, such as DIF)
Questionnaire development (for students, teachers, parents, schools)	No national agency prepares supplementary questionnaires	A national agency prepares questionnaires with technical support from another national or international agency	A national agency prepares questionnaires independently and field tests them
B. Administrative Capacity			
Project coordination	No national agency has any experience in managing school-based data collection	A national agency regularly conducts assessments of student achievement	A national agency regularly contracts with other agencies that have the expertise to carry out the various tasks associated with an assessment of student achievement
Data collection and test administration	No national agency regularly collects data	A national agency collects achievement data from students following detailed instructions prepared by another national or international agency	A national agency prepares administration manuals for data collection and exception reporting, and regularly collects student achievement data
Data entry and quality control	No national agency regularly enters data	A national agency enters data following directions from another national or international agency, including using data entry programs (such as WinDEM), with assistance	A national agency prepares or adapts software for data entry and quality control, and routinely uses it for creating computer data files
Professional and reliable scoring of constructed response items	No national agency scores achievement tests, or constructed response items are not scored reliably, or tests are scored by an international agency	A national agency professionally scores achievement test items using clear scoring rubrics that were developed by another national or international agency; items are scored reliably	A national agency routinely develops clear scoring rubrics and applies them to professionally score constructed response items; items are scored reliably
C. Analytic Capacity			
Sampling schools, classrooms or students (as	No national agency has a comprehensive list of schools from	With assistance, an national agency develops a sampling frame and draws a national sample	A national agency independently draws a national sample of schools based on statistical principles and

Capacity Dimension	Lowest Capacity	Modest Capacity	Highest Capacity
needed)	which to develop sampling frame	of schools and classrooms (students), using procedures or software developed by experts from another agency	uses this sample for national assessment
Weighting (as needed)	No national agency regularly weights data	A national agency weights data, following directions from another national or international agency	A national agency prepares or adapts software for weighting and routinely uses it
Scaling	Results on tests are reported using the number or percent correct of items for each student, or using scale-like scores created without using modern or classical psychometric methods	A national agency uses psychometric methods to create scales from scores (e.g. IRT methods), with assistance from another national or international agency	A national agency independently uses modern psychometric methods such as IRT scaling to create achievement scales.
Equating	Assessments are not comparable over time, and may differ in content, level of difficulty, and target population of students	Some effort is made to make the assessments measure the same competencies over time	Assessments are statistically equated over time and can measure change
Technical report writing	No technical reports are written	A national agency writes technical reports with assistance, following detailed models provided by another national or international agency	A national agency independently designs, writes and produces technical reports that meet professional standards, including presentation of standard errors, non-reporting of non-significant differences
Measurement training	No national university has departments of statistics or psychometrics	University departments of statistics and/or psychometrics award master's degrees	University departments of statistics and/or psychometrics award internationally-recognized doctoral degrees
D. Communication Capacity			
Policy report writing	No reports are written	A national agency writes policy reports with assistance, following detailed models provided by another national or international agency	A national agency writes policy reports independently and prepares clear and succinct policy notes for the Ministry of Education
Communication of findings from assessment	No national agency communicates the results of the assessments to the media or government	A national agency communicates results following detailed model provided by another national or international agency	A national agency prepares reports, develops its own strategy for communication of results, including the use of print and electronic media, and organizes meetings with policy makers to explain results
Utilization of results for decision making	No actions based on assessment results are taken or only token actions are announced (but not implemented)	A series of actions are initiated in response to assessment results	Assessment results are used frequently and accurately to justify reforms in education policy and practice

*A national agency may be one of any number of organizations within a country: the Ministry of Education, a department within the Ministry of Education, an independent testing or research organization, a university or university department. Different national agencies may be responsible for different aspects of assessment, and the term should not be interpreted as meaning only one agency.

Annex E: Low- and Middle-Income Countries with Regularly Conducted National Assessments and their Participation in International and Regional Assessments

Country with Regular NAS	Assessment	IEA	OECD	SACMEQ	LLECE	PASEC
Albania	Yearly					
Argentina	Yearly, periodic	√	√		√	
Belize	Yearly	√				
Brazil	Yearly		√			
Chile	Yearly	√	√			
Colombia	Yearly	√	√		√	
Costa Rica	Yearly				√	
Cuba	Periodic				√	
Dominican Republic	Yearly				√	
El Salvador	Yearly	√			√	
Estonia	Yearly	√	√			
Gambia	Yearly					
Ghana	Yearly, periodic	√				
Guatemala	Yearly				√	
Guyana	Yearly					
Honduras	Periodic	√			√	
Hungary	Yearly, periodic	√	√			
India	Periodic	√				
Indonesia	Yearly	√	√			
Jamaica	Yearly					
Jordan	Yearly	√	√			
Malawi	Periodic			√		
Malaysia	Yearly	√				
Mauritania	Yearly					
Mexico	Yearly	√	√		√	
Mongolia	Yearly	√				
Morocco	Periodic	√				
Peru	Periodic		√		√	
Poland	Yearly	√	√			
Seychelles	Yearly					
South Africa	Yearly	√		√		
Thailand	Yearly	√	√			
Turkey	Yearly	√	√			
Uruguay	Periodic		√		√	
<i>Percent</i>		.59	.41	.06	.32	

Source: UNESCO 2007 Annex: National Learning Assessments; IEA, OECD, UNESCO, CONFEMEN

Annex F. Institutional representatives of low and middle income countries at 2007 IEA General Assembly Meeting, Hong Kong

Low and Middle Income country attending 2007 IEA GA meeting	Institution represented	Same institution for NLA in GMR 2008?
Argentina	MOE Direccion Nacional de Informacion y Evaluacion de la Calidad Educative	yes
Botswana	MOE Examinations and Research and Testing divisions	--
Brazil	MOE National Institute for Educational Studies and Research (INEP)	yes
Bulgaria	University of Sofia, Faculty of Mathematics and Informatics	no
Chile	MOE Unit of Curriculum and Assessment	yes
China	China National Institute for Educational Research	--
Czeck Republic	Institute for Information on Educaton	--
Egypt	National Center for Examinations and Educational Evaluation (of MOE)	yes
Estonia	Tallinn University	no
Georgia	National Assessment and Examinations Center	yes
Hungary	MOE	no
Indonesia	MOE National Office for Educational Research and Development	no
Iran	MOE Institute for Educational Research	--
Jordan	National Center for Human Resources Development	no
Kazakhstan	National Centre for Assessment of the Quality of Education	--
Latvia	University of Latvia, Faculty of Education and Psychology	--
Lithuania	MOE National Examination Center	yes
Malaysia	MOE Educational Planning and Research Division	no
Mexico	Instituto Nacional para la Evaluacion de la Educacion (INEE)	yes
Morocco	MOE Direction de l'Evaluation and de l'organisation de la vie scolaire	yes
Palestinian Authority	MOE Assessment and Evaluation Center	--
Philippines	Science Education Institute, Department of Science and Technology	no
Romania	Institute of Educational Sciences	no
Russian Fed.	Russian Academy of Education, Center for Evaluation of the Quality of Education	--
Slovak Republic	National Institute for Education	yes
Slovenia	University of Ljubljana, Education Research Institute	--
South Africa	Human Sciences Research Council	yes
Thailand	Office of the Education Council	no
Turkey	MOE Education Research and Development Directorate	yes