International Assessment Surveys of Educational Achievement in Developing Countries. Why Education Economists should Care.

Gérard Lassibille

International Assessment Surveys of Educational Achievement in Developing Countries: Why Education Economists Should Care

GÉRARD LASSIBILLE

Centre National de la Recherche Scientifique and Institut de Recherche sur l'Economie de l'Education, Pôle AAFE, Esplanade Erasme, BP26513-21065 Dijon Cedex, France.

Published in Comparative Economic Studies, 2015, 57 (4), pp. 655-668
doi:10.1057/ces.2015.19

This paper reviews the most known international assessment studies that are conducted in the context of poor countries and highlights the lack of empirical evidence on the degree to which the contents of the tests really match countries' curricula. To illustrate, the paper evaluates the sensitivity of an international testing instrument by comparing the responses of students in two consecutive grades on the same battery of tests. Using propensity score matching to control for student and teacher characteristics, the results show that the tests are not grade sensitive, which raises the question of the validity of many empirical works that are based on similar instruments.

Keywords: international achievement tests, content validity, developing countries

JEL Classification: I21, P52

INTRODUCTION

Education economists have devoted considerable empirical attention to the performance of the education sectors (see, eg, Hanushek, 1986, 1995; Pritchett and Filmer, 1999). Student test scores from a battery of international tests are the cornerstone of their analysis of student achievement that aim to assess the efficiency of schooling systems and to generate recommendations for policymaking in order to improve the quality of education. These assessment surveys of educational achievement raise a large number of conceptual, methodological and practical issues that have been widely documented in the literature (Greany and Kellaghan, 2008; Winter, 1998). One aspect to which users of international assessment studies probably do not pay quite enough attention is the quality of the testing instruments, and their ability to adequately measure students' performance. While education economists who study student performance have paid attention to the quantification of inputs and, more recently, to the specification of their empirical models, they have placed little emphasis on the quality of the achievement test scores they use to evaluate the outcomes. Obviously, addressing these quality concerns is vitally crucial for those who want to properly inform decision making in education and to address policy issues.

This paper questions the ability of these instruments to measure student performance accurately. A review of the main assessment studies conducted in the context of low-income countries indicates that no technical documentation is provided that demonstrates that the contents of the tests match the curricula. For this reason, it is

generally extremely difficult to judge the quality of the assessment data. Taking advantage of data of uncommon richness, this paper evaluates, by way of example, the content validity of the Programme d'Analyse des Systèmes Educatif de la Confemen (PASEC) conducted in Madagascar. The results are based on a cohort of around 15,000 grade-three and grade-four students who were enrolled in a total of about 1,000 public primary schools and who were tested on the same grade-four teaching program, whatever grade they were enrolled in. By comparing the scores achieved by students in grades three and four on the same grade-four test, the available data make it possible to measure *ex post* the sensitivity of the test instruments and to offer useful lessons on the way some international achievement tests are calibrated. The paper uses propensity score matching to control for student and teacher characteristics. In the absence of being able to carefully compare the test items with the curriculum content in Madagascar, the results show that the PASEC instruments for this country are probably poorly calibrated and do not properly allow meaningful analysis of the performance of the education sector.

Although the findings relate to a particular country, they raise important questions about the reliability of the available battery of tests for use in developing countries, and the accuracy of the policy recommendations that can be derived from this kind of assessment data. The results may be suggestive for education economists who use them for their analyses and may be useful for education specialists who design these instruments. The paper adds to the empirical literature on the curricular coverage of international testing efforts. While test-curriculum matching analysis has been conducted to investigate the appropriateness of some 'gold standard' tests like the Third International Mathematics and Science Study (TIMSS; see, eg, Beaton and Gonzalez, 1997), this paper reports on the first known attempt to assess *ex post* the accuracy of a battery of tests designed for use in the context of developing countries.

The remainder of the paper proceeds as follows. The next section briefly reviews the main international assessment surveys that are conducted in developing countries. The section after that describes the data; the following section explains the estimation strategy and presents the results; and the final section concludes.

AN OVERVIEW OF THE MAIN INTERNATIONAL ASSESSMENT STUDIES CONDUCTED IN DEVELOPING COUNTRIES

The vast majority of poor countries do not take part in the most known international studies of achievement like the Progress in International Reading Literacy Study (PIRLS), the Programme for International Student Assessment (PISA) or the Third International Mathematics and Science Study (TIMSS). In these countries, regional assessment studies have been carried out since the beginning of the 1990s in order to provide a basis for the evaluation of educational policy and practices. This section briefly reviews the most known international surveys conducted in developing countries, focusing on the lack of information to be able to judge really the quality of the assessment data.

The Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE), coordinated by the UNESCO Regional Office for Latin America and the Caribbean (OREALC), provides information on student learning since 1998. The aim of the project is to inform the formulation and the impact of education policies at primary

education level within Latin American countries. Sixteen countries participated in the last project conducted in 2002 by the LLECE (OREALC, 2008). Around 200,000 students from 3,000 schools in the region were tested in reading, writing, mathematics and science. Although the designers of the program indicate that the test instruments are closely matched to the curricula, no documentation is provided to demonstrate empirically the content validity of the battery of tests.

The Programme d'Analyse des Systèmes Éducatifs de la CONFEMEN (PASEC) is conducted under the auspices of the Conférence des Ministres de l'Éducation des Pays ayant le Français en Partage (CONFEMEN). It was launched in 1991 at a conference of French-speaking education ministers held in Djibouti. The first study was carried out in 1993. To date, 18 French-speaking African countries have participated in the PASEC program. Grade-two and grade-five students enrolled in both public and private primary schools are tested in French, mathematics and national language. Unlike many others assessment studies, students are tested at the beginning and end of the school year, making it possible to model the determinants of learning as being a dynamic process over the course of an academic year. The PASEC instruments are theoretically designed to evaluate the extent to which curriculum objectives have been met, and to measure what students in a particular grade have learned. However, no comprehensive documentation is provided that demonstrates curriculum match between country curricula and the test content. Nevertheless, based on the available information, many national reports are periodically published by the PASEC team members, with the objective to inform decision making in education and to address policy issues in French-speaking African countries.

The Joint UNESCO-UNICEF Montoring Learning Achievement (MLA) project started in the early 1990s as a result to the Jomtien World Conference on Education for All. It is designed to identify appropriate policies for improving the quality of basic education, aiming to assist countries in building and strengthening national capacities, and providing policymakers with analytical tools to monitor the quality of their basic educational programs (Chinapha, 1995). Seventy-two countries, all over the world, have participated in the MLA program since 1992. The MLA instruments are intended to measure basic learning outcomes in literacy, numeracy and life skills at grades four and five, and mathematics, science and life skills at grade eight. Like in other assessment projects, data on students, schools and home background conditions that are considered relevant to student achievement are also collected. A search of published and publicly available reports contained in the MLA website indicates that the test construction processes are poorly documented. Comprehensive documentation of content validity evidence for the MLA instruments is apparently not available, and it is extremely difficult to appreciate the quality of the assessment data.

The Southern Africa Consortium for Measuring Education Quality (SACMEQ) was launched in 1995 with the assistance of the International Institute for Educational Planning (IIEP) of UNESCO. The program is a network of ministries of education in southern and eastern Africa, comprising 15 Anglophone countries (IIEP, 2006). SACMEQ assesses grade-six students in reading and mathematics. One of the objectives of the program is to compare changes in educational quality and educational achievement over time. A large number of country reports have been produced or are in preparation, each of which identifying factors that might account for the variation in student achievement levels in order to help inform policy making. Many technical

reports describing the design of the assessments, the methods of sampling, the implementation of the assessment data have been published, and are easily accessible to the researchers and the users of the data. However, as noted by an external evaluation team recruited by UNESCO to review the quality of the SACMEQ program (see Ercikan *et al.* , 2008), no documentation is provided that demonstrates curriculum match between countries' curricula and the test content even though curricular match considerations are included in the test construction process. In this regard, the evaluators recommended to SACMEQ to develop and make publicly available comprehensive documentation of content validity evidence for the test instruments and to revise periodically this content validity evidence.

The evidence that emerges from this review shows that the international assessment surveys do not provide convincing evidence for construct validity of the tests. The technical documentation generally indicates that the tests instruments fit with the contents of the curricula and textbooks used in the participating countries. This kind of statement is supported to some degree by the fact that items are generally developed by representatives of each country or are sent to countries for review by their curriculum specialists. However, no technical documentation is provided that demonstrates that the contents of the tests match the countries' curricula, even though content validity evidence is essential for judging the quality of the assessment data and is a major requirement according to the professional testing guidelines developed jointly by the American Educational Research Association (AERA), American Psychological Association (APA) and the National Council on Measurement in Education (NCME) (see AERA *et al.* , 1999).

THE DATA

Data for the analysis come from surveys fielded with World Bank support in the framework of a school management program implemented on an experimental basis by the Ministry of Education of Madagascar (the Amélioration de la Gestion de l'Education à Madagascar - AGEMAD). Student achievement tests were administered in June 2007 to a cohort of pupils who were enrolled in grade three in the 2005-2006 school year and who were tested on the grade-four teaching program, whatever grade they were enrolled in during the 2006-2007 school year. Moreover, in February 2006 all the students were tested on the grade-three teaching program. By comparing the scores achieved by students in grades three and four on the same grade-four test, the available data make it possible to measure *ex post* the sensitivity of the test instruments and to offer useful lessons about the way some international achievement tests are calibrated.

Table 1: Number of items administered to students in grades three and four

Subject	Original PASEC instruments (2004–2005)	PASEC items tested (2006–2007)
Math	34	28
French	43	27
Malagasy	38	25
Total	115	80

Source: 2007 AGEMAD test

The test instruments used items from the pre-test administered in Madagascar to grade-five pupils by the PASEC in September 2004. Because the curricula in primary education have changed slightly since 2004, items not exactly aligned with the new programs were excluded. A total of 80 items were thus selected from the 115 test items included in the 2004-2005 PASEC database: 27 in French, 25 in Malagasy and 28 in mathematics (Table 1). The selection was carried out by the learning and curriculum division of the ministry of education. The items from the PASEC test were administered in June 2007 to 11,611 pupils in grade four and to 4,379 pupils in grade three, in a total of about 1,000 public schools. [1] Following the PASEC methodology, a maximum of 25 pupils per school were randomly selected and tested in the three subjects[2]; it is noteworthy that every school in the sample had students tested in grades three and four. The test was conducted under the direct supervision of the Institut National de la Formation Pédagogique and the PASEC Malagasy team; it was administered following standard practices in such studies (see World Bank, 2010). All administrators followed the same testing procedures and gave the same instructions to all examinees; no irregularities occurred during the administration and grading of the test. Like many other testing instruments, the PASEC test is a multiple-choice test. Each item has four distractors on average; otherwise stated, the probability of achieving a high score by random guessing (ie, without knowledge) is small. As explained above, the test is theoretically designed to evaluate the extent to which curriculum objectives have been met in an educational program and to measure what students in a particular grade have learned.

THE RESULTS

This paper uses propensity score matching as a methodology to compare the scores achieved by students in grades three and four on the grade-four test that was administered to both categories of students at the end of the 2006-2007 school year. The propensity score matching technique seeks to ensure that grade-four students (treatment group) and grade-three students (control group) differ only in that the former group was exposed to the grade-four program and the latter group was not. The propensity score (Rosenbaum and Rubin, 1985) is defined as the conditional probability of being exposed to the grade-four program given a set of observed covariates or pre-treatment characteristics. Following the common practice, a probit model is used to estimate the propensity score. The prediction of group membership is based on the following individual and family background variables: student's age and gender, mother's and father's education level, and household wealth proxied by whether the household is equipped with electricity. Students still enrolled in grade three in 2006-2007 differ in individual ability from the rest of the cohort. To account for the differences in ability, I match for ability using previous academic performance, as reflected in the percentage of correct responses to the grade-three tests administered to the cohort of students in the 2005-2006 school year. [3] Because students' test scores may be influenced by teacher characteristics, the type of employment contract held by teachers is also included among the set of covariates.[4] On the basis of their job status, there are two broad categories of teachers in Madagascar. The first group consists of civil service teachers who are recruited and paid by the government. The second group consists of contract teachers. In Madagascar, as well as in many other developing countries, teachers' employment status is a good proxy for teachers' qualifications. While contract teachers in Madagascar have more years of formal schooling than civil service teachers on average, they have little prior exposure to teacher training programs, and their professional

qualifications are significantly lower (World Bank, 2010). Appendix Table A1 presents probit estimates of the coefficients used for propensity score matching. For the analysis presented here, the weighted nearest-neighbor matching method is used to define the comparison group of students for the sample. [5] Under this method, each treated student is matched to the control student with the closest propensity score; to improve the quality of matches the analysis is restricted to the region of common support. By matching on the single propensity score, it is possible to form a control group that is quite similar to the treatment group, as shown in Table 2.

Table 2: Characteristics of unmatched and matched samples of students

Variable	Sample	Mean		% bias	% reduction |bias|	t-test	
		Grade-four students (treatment group)	Grade-three students (control group)			t	p>t
Students' characteristics							
Score in grade three	Unmatched	63.982	53.725	64.1		36.28	0.000
	Matched	64.010	63.951	0.4	99.4	0.29	0.774
Age (in years)	Unmatched	11.590	11.267	21.0		11.43	0.000
	Matched	11.586	11.555	2.0	90.6	1.44	0.151
Girl (%)	Unmatched	48.670	44.211	8.9		4.95	0.000
	Matched	48.785	46.926	3.7	58.3	2.74	0.006
Mother literate (%)	Unmatched	90.325	87.732	8.3		4.66	0.000
	Matched	90.584	92.238	5.3	36.2	−4.09-	0.000
Father literate (%)	Unmatched	87.690	85.267	7.1		3.78	0.000
	Matched	87.712	87.309	1.2	83.4	0.86	0.392
Household equipped with electricity (%)	Unmatched	5.240	4.244	4.7		2.54	0.011
	Matched	5.268	5.056	1.0	78.7	0.71	0.481
Class-level variable							
Civil servant teacher (%)	Unmatched	51.174	50.947	1.3		0.75	0.451
	Matched	51.579	50.925	1.3	−4.3	1.07	0.286

Source: 2007 AGEMAD test

Because students in grade three were not exposed to the grade-four program, they are expected to achieve a low score on the grade-four test. What does the evidence reveal in this regard? Table 3 presents the summary statistics for the matched sample of students, with the means of the percentage of correct answers on the tests for students in each grade. Several features of the results are noteworthy. As expected, students in grade three perform worse than pupils in grade four who took the same tests, on average. However, differences between each group of students are small. In mathematics and Malagasy, the scores of fourth graders exceed the scores of third graders by 6% and 7%, respectively; in French, students in grade four have only a 2 percentage point advantage over students in grade three. [6] The results based on the combined scores in the three subjects indicate that fourth graders only have a 5 percentage point advantage, on average, over third graders.

Table 3: Correct answers to grade-four test

Grade level tested	Correct answers (%)			
	Math	French	Malagasy	Total
Pupils in grade three	47.4	27.2	41.5	38.7
Pupils in grade four	54.7	29.2	48.0	43.9
Difference between grades four and three	7.29*	2.0*	6.4*	5.24*

Source: 2007 AGEMAD test
** = Significant at 1%.*

Figure 1 better illustrates the differences in test scores by showing the distribution of the percentage of correct answers in the three subjects for the two matched samples of students. The figure clearly indicates that the distributions have considerable overlap and that the test instruments are probably not very grade sensitive (see our discussion below). In mathematics and Malagasy, around 85% of the two distributions overlap. The overlapping is even larger in French, with 89% of the two distributions overlapping. [7]

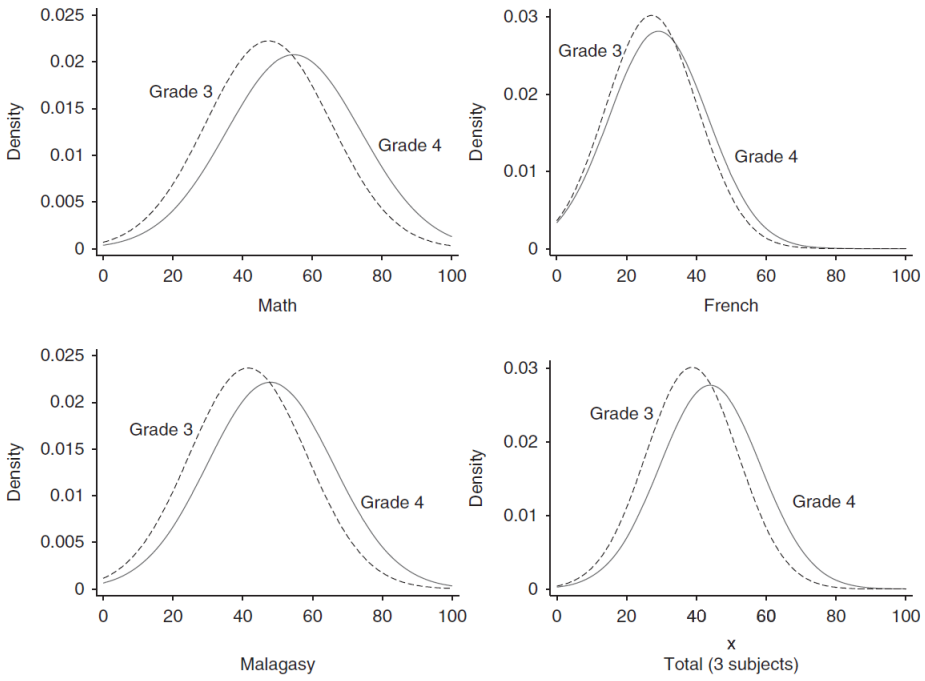

Figure 1*:* Distribution of % correct answers, by grade and subject.
Source: 2007 AGEMAD post-test

What can explain these intriguing results? Each grade-four student was matched with a grade-three student with the same propensity score. As shown in Table 2, the two groups of students were matched quite well and are virtually identical. In other words, no apparent random student or teacher effects can explain the results. [8] One possible explanation for the high score of grade-three students is that, although third graders did not know the answers to the items, they were lucky in guessing correct answers. However, that is not very plausible. As explained above, the number of distractors for each item in the multiple-choice test is large, and for this reason the probability of gaining a high score without knowledge is small. Differences in the availability of

pedagogical and physical resources that are directly related to classroom teaching could explain part of the results documented above. However, this hypothesis is unlikely. As shown in Table 4, third and fourth graders in the matched sample are enrolled in schools with similar characteristics. In both sub-samples, schools operate, on average, with identical pupil-teacher ratios. In terms of the availability of pedagogical resources, no significant differences exist between third and fourth graders. The quality of the physical facilities of the schools is similar in each sub-sample of students. [9]

Another possible explanation for the high scores of grade-three students might be that instruction in grade four is not aligned with the grade-four program. However, there is no objective reason to believe that fourth grade teachers are less likely than their counterparts to follow the program strictly. Moreover, sub-district officers inspect schools frequently in Madagascar to ensure that education programs are effectively carried out. According to the AGEMAD school survey, sub-district officers visited 90% of the public primary schools in Madagascar during the year previous to the survey; about 75% of inspections were concerned with pedagogical matters.

The most likely reason why students in grade three perform so well on the grade-four test is that the test is not adequately calibrated and does not provide a valid assessment of knowledge in relation to the program objectives. Otherwise stated, the empirical evidence suggests that the test is imperfectly aligned with the intentions of the grade-four curriculum and probably fails to recognize learning that has really occurred at this stage of primary education. Of course, one might object to this conclusion by pointing out that (a) some material is common to both third and fourth graders and for this reason the grade-four test may include some third grade material and (b) third graders - at least the most advanced students - may know some of the fourth grade material. While these arguments certainly explain why some students in grade three may perform as well as - or better than - other students in grade four, they probably cannot justify why the overlap between the distribution of the two samples of students is so considerable (up to 80%).

CONCLUSION

The vast majority of developing countries are not included in 'gold standard' tests like PIRLS, PISA or TIMSS. In these countries, specific evaluation systems have been carried out since the beginning of the 1990s. This paper reviewed the most known assessment surveys that are conducted in the context of developing countries. This review has shown that the test construction processes are poorly documented. No empirical evidence or proof is provided to confirm that the tests provide adequate curriculum coverage. At best, the assessment framework documents only inform test users that the test items are aligned with the curricula, without providing any empirical evidence on the degree to which test content matches country curricula.

Taking advantage of data of an uncommon richness, this paper checked *ex post* the accuracy of an international testing instrument designed for use in the context of developing countries. To do so, propensity score matching was used to compare the results obtained by students in grades three and four who took the same grade-four test at the end of the school year. The results showed that up to 89% of the distribution of the percentage of correct answers for the two samples of students overlap. As discussed above, it is unlikely that this large overlap is attributable to significant differences

between the two samples across: (a) students and teachers characteristics that determine students' performance; (b) the probabilities of gaining a high score without knowledge; (c) the availability of pedagogical and physical resources that are directly related to classroom teaching; (d) the probability for teachers to follow the programs strictly. The most likely reason why so many third graders do so well on a test designed for fourth graders is that there is probably an imperfect correspondence between what is taught in the fourth grade and what is covered by the test.

The results of this study have been discussed with the Malagasy team responsible for constructing the PASEC test. One remarkable and positive aspect of that discussion has been recognition that the test is not aligned with the curricula and needs serious adjustments. This conclusion raises serious questions about the findings of many empirical works that investigate the efficiency of the education sector using this battery of tests and casts doubt on the reliability of the policy recommendations that this kind of work can offer. Because the results are based on a particular developing country, they are not generalizable to other contexts or to other international testing instruments. However, two important lessons of this paper are that education economists who use international achievement tests for their analysis should be careful when interpreting their results and that test developers should take care to ensure that a good match exists between the curricula's intentions and the tests. To avoid doubts and errors, alignment studies should be systematically carried out, and the results should be made available to the potential users of international batteries of tests.

Acknowledgements

The author is grateful to the World Bank and the Ministry of Education of Madagascar for providing access to the AGEMAD (Amélioration de la Gestion de l'Education à Madagascar) database used here. Funding from the European Union and the Government of Andalusía through grants P09SEJ4859 and FC-14-ECO-15 is gratefully acknowledged. The opinions expressed in this paper are those of the author alone and should not be attributed to the institution with which he is associated, to the World Bank or to the governments listed above. For helpful discussions and comments, the author thanks Per-Erik Lyren, Murielle Nicot-Guillorel, Lina Rajonhson, and participants at both the Fourth Workshop on Improving Education Management in African Countries and the 13th SweSAT Conference.

Footnote

[1] Following common practices in international assessment surveys, the sample design consisted of a two-stage design. In a first stage, a sample of schools stratified by school size and representing all geographical areas of the country was selected. In a second stage, a simple random sample of a fixed number of pupils within selected schools was selected. The original database included about 300 schools that were implementing new pedagogical approaches when the test was administered. For the purpose of the analysis, and in order to obtain a homogeneous group, these schools were dropped from the sample.

[2] In schools with less than 25 pupils in grade three or four, all the students were tested.

[3] Students were tested in mathematics, French and Malagasy. Individual ability is proxied by the percentage of correct responses in these three subjects.

[4] Because of data limitations, it is not possible to control the matching process for teacher experience. However, this omission probably does not affect the results presented here, as previous research has shown that teacher experience has no significant impact on students' performance in Madagascar (Lassibille and Tan, 2003).

[5] Different matching methods were tested; the nearest-neighbor matching technique produced the most balanced sample.

[6] If the analysis had not used propensity score matching and had just compared grade-three and grade-four students, the results would have been different.

[7] Because testing efforts attempt to cover a broad range of material, the test score distributions almost always overlap. However, the main point is that the overlap is considerable.

[8] In other words, the large overlap between the distribution of the two samples of students cannot be attributed to significant differences in the most important factors that determine students' performance.

[9] This index is constructed using principal components analysis based on the following physical conditions of the school: the structure is permanent, the number of classrooms is sufficient, and the school is equipped with electricity, water, latrines and chairs for all pupils. The index ranges from about 126 in schools with all of these features to 75 in schools with none of them.

References

AERA (American Educational Research Association), APA (American Psychological Association) and NCME (National Council of Measurement in Education). 1999: *Standards for educational and psychological testing*. AERA: Washington DC.

Beaton, A and Gonzalez, E. 1997: TIMSS test-curriculum matching analysis. In: Martin, M and Kelly, D (eds). *Third International Mathematics and Science Study - Technical report*. Boston College: Chestnut Hill, MA.

Chinapha, V. 1995: *The monitoring project moving ahead*. Studies and Working Document no.9, UNESCO, Division of Basic Education: Paris.

Ercikan, K, Arim, R, Oliveri, M and Sandilands, D. 2008: *Evaluation of dimensions of the work of the Southern and Eastern Africa consortium for monitoring educational quality (SACMEQ) and of its programme of cooperation with the international institute for educational planning (IIEP)*. UNESCO: Paris.

Greaney, V and Kellaghan, T. 2008: *Assessing national achievement levels in education*. The World Bank: Washington DC.

Hanushek, E. 1986: The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature* **49** (3): 1141-1177.

Hanushek, E. 1995: Interpreting recent research on schooling in developing countries. *World Bank Research Observer* **10** (2): 227-246.

IIEP (International Institute for Educational Planning). 2006: *Lettre d'information de l'IIPE no.1*. IIEP: Paris.

Lassibille, G and Tan, JP. 2003: Student learning in public and private primary schools in Madagascar. *Economic Development and Cultural Change* **51** (3): 699-717.

OREALC (Oficina Regional de Educación para América Latina y el Caribe). 2008: *Primer Reporte de Resultados del Segundo Estudio Regional Comparativo y Explicativo (SERCE)*. UNESCO: Santiago.

Pritchett, L and Filmer, D. 1999: What education production functions really show: A positive theory of education expenditures. *Economics of Education Review* **18** (3): 223-239.

Rosenbaum, PR and Rubin, DB. 1985: Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician* **39** (1): 33-38.

Winter, SJ. 1998: International comparisons of student achievement: Can they tell us which nations perform best and which education systems are the most successful? *Education 3 to 13* **26** (2): 26-32.

World Bank. 2010: *Améliorer la gestion de l'enseignement primaire à Madagascar: résultats d'une expérimentation randomisée*. The World Bank: Washington DC.

Appendix

Table A1: Propensity function parameter estimates: Probit on student enrollment in grade four

Indicator	Coefficient	z-statistic
Students' characteristics		
Score in grade three	0.0233	32.44***
Age (in years)	0.0759	10.13***
Girl	0.1442	6.31***
Mother literate	0.1119	3.00**
Father literate	0.1199	3.50***
Household equipped with electricity	0.0788	1.67*
Class-level variable		
Civil servant teacher	0.0384	1.97**
Constant	−1.9134	−17.69***
Number of observations	14,988	
LR $\chi 2$	1,345.80***	
Pseudo R2	0.0764	

Source: 2007 AGEMAD test
*** = significant at 1%; ** = significant at 5%; * = significant at 10%