

**A Millennium Learning Goal:  
Measuring Real Progress in Education**  
By Deon Filmer, Amer Hasan and Lant Pritchett

**Abstract**

The Millennium Development Goal for primary schooling completion has focused attention on a measurable *output* indicator to monitor increases in schooling in poor countries. We argue the next step, which moves towards the even more important Millennium Learning Goal, is to monitor *outcomes* of learning achievement. We demonstrate that even in countries meeting the MDG of primary completion, the majority of youth are not reaching even *minimal* competency levels, let alone the competencies demanded in a globalized environment. Even though Brazil is on track to meet the MDG, our estimates are that 78 percent of Brazilian youth lack even minimally adequate competencies in mathematics and 96 percent do not reach what we posit as a reasonable global standard of adequacy. Mexico has reached the MDG—but 50 percent of youth are not minimally competent in math and 91 percent do not reach a global standard. While nearly all countries' education systems are expanding *quantitatively* nearly all are failing in their fundamental purpose. Policymakers, educators and citizens need to focus on the real target of schooling: adequately equipping their nation's youth for full participation as adults in economic, political and social roles. A goal of school completion alone is an increasingly inadequate guide for action. With a Millennium Learning Goal, progress of the education system will be judged on the *outcomes* of the system: the assessed mastery of the desired competencies of an entire age cohort—both those in school and out of school. By focusing on the learning achievement of all children in a cohort an MLG eliminates the false dichotomy between “access/enrollment” and “quality of those in school”: reaching an MLG depends on both.

---

The Center for Global Development is an independent think tank that works to reduce global poverty and inequality through rigorous research and active engagement with the policy community.

Use and dissemination of this Working Paper is encouraged, however reproduced copies may not be used for commercial purposes. Further usage is permitted under the terms of the Creative Commons License. The views expressed in this paper are those of the author and should not be attributed to the directors or funders of the Center for Global Development.

**A Millennium Learning Goal:  
Measuring Real Progress in Education**

Deon Filmer<sup>1</sup>, Amer Hasan<sup>2</sup> and Lant Pritchett<sup>3</sup>

Center for Global Development and The World Bank

June 22, 2006

---

<sup>1</sup> Senior Economist, Development Economics, World Bank.

<sup>2</sup> Researcher, Center for Global Development and Development Economics, World Bank.

<sup>3</sup> Lead Socio-Economist, South Asia Region, World Bank and Non-Resident Fellow, Center for Global Development.

## A Millennium Learning Goal: Measuring Real Progress in Education

### Introduction

The United Nations' Millennium Development Goals (MDGs) seek to “[e]nsure that by 2015, children everywhere, boys and girls alike will be able to complete a full course of primary schooling.” Progress towards this goal is typically measured by the three targets: the net enrollment ratio in primary education<sup>4</sup>; the proportion of children who complete the primary school cycle<sup>5</sup>; the literacy rate of 15-24 year olds<sup>6</sup>. In addition, an MDG for gender parity is measured by the ratio of female to male enrollment in the primary and secondary school cycles. The World Bank, among others, has favored the primary completion rate as the indicator that best reflects the MDG education goal that children “complete a full course of primary schooling.” By this indicator, the world has made substantial progress. The primary completion rate in low-income countries increased from 66 to 74 percent between 1991 and 2004, with growth in all of the poorer regions: Latin America and the Caribbean (86 to 97 percent); Middle East and North Africa (78 to 88 percent); South Asia (73 to 82 percent); and Sub-Saharan Africa (51 to 62 percent).<sup>7</sup>

But universal completion of primary *school* has always been only a means to the actual goal of universal *education*: that every youth should make the transition to adulthood equipped with the minimal set of competencies—including both cognitive and non-cognitive skills—

---

<sup>4</sup> Net enrollment ratio is defined as total enrollment of primary school age children divided by the population of primary school age children.

<sup>5</sup> The primary completion rate is typically measured by its proxy: the ratio of the number of non-repeaters in the terminal grade of primary school to the number of children of the official age of the terminal year of primary school.

<sup>6</sup> Literacy is defined by UNESCO as the percentage of 15-24 year olds who can, with understanding, both read and write a short simple statement on their everyday life—but whether this is measured with any accuracy is far from certain. Many countries report literacy numbers with widely varying definitions, some as rudimentary as being able to sign one's name.

<sup>7</sup> <http://www.developmentgoals.org>.

needed to function adequately in the economic, social, and political spheres of a modern society.<sup>8</sup> The recent *World Development Report 2007: Development and the Next Generation* builds on this notion of childhood and youth as a time to prepare for transitions and the critical role of schooling as not about rote recitation or mastering facts, but improving the skills of young people for work and life—making education opportunities more relevant to the needs of young people as future workers, parents and citizens.<sup>9</sup> Why is the goal that children universally complete a primary cycle, usually of 5 or 6 years? Why not 2 years? Why not 14 years? The underlying rationale for *schooling* goals has always been broad *learning* goals. Education specialists typically had in mind a set of minimally adequate knowledge, skills, attitudes, values, behaviors, which we broadly call “competencies”<sup>10</sup>, to be acquired through schooling. The duration and curriculum of primary (or “basic”) schooling were then set so that completion of the cycle with at least some mastery of the curriculum implied acquisition of the universally necessary competencies.

The learning profile—the relationship between competencies and years of schooling—links the *output* goal of universal completion (the education MDG) with the *outcome* goal of universal competencies (MLG). The implicit assumption in the MDG is that the learning profile is sufficiently steep that the average, and even low performing, students reach the threshold on

---

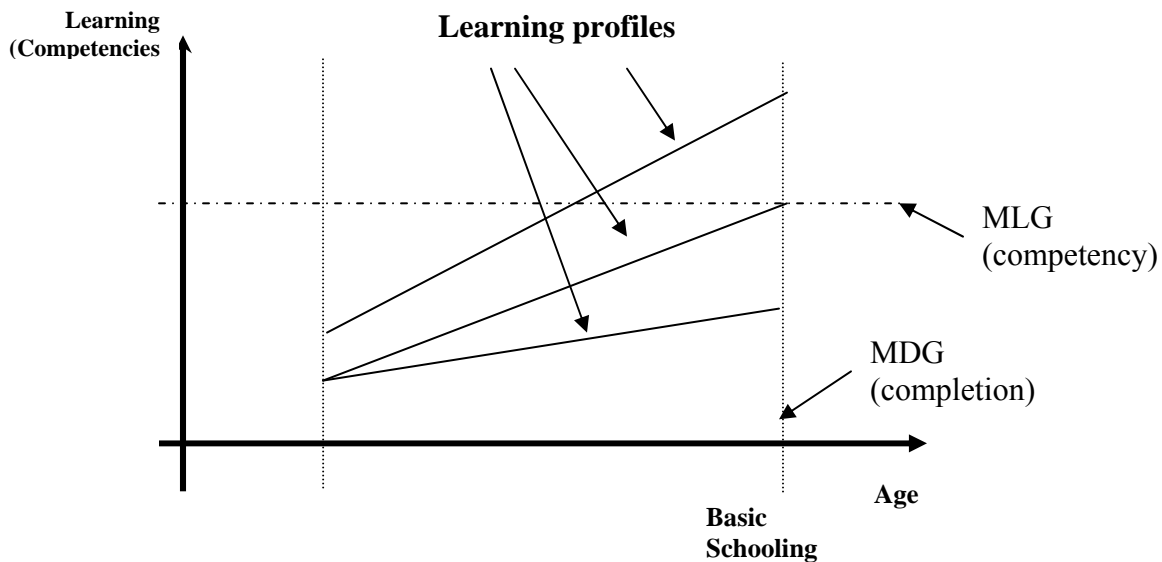
<sup>8</sup> This normative societal “should” has been rationalized in various ways: arguments that are “rights”, “equity” or “fairness” based, or through arguments for pragmatic social outcomes.

<sup>9</sup> World Bank (2006), see in particular Chapter 3, “Learning for Work and Life”. The WDR2007 documents dramatic shortfalls in skills among youth in developing countries, and suggests policies to expand the educational opportunities available to young people, enhance young people’s abilities to fully take advantage of the opportunities they face and choose the educational path most suited to them, and to ensure that youth who never went to school or dropped out before completing primary school have a second chance at acquiring basic skills. All of these policies, however, are geared towards preparing youth for the various transitions to adulthood they will ultimately make.

<sup>10</sup> This is not a debate about narrow skill based versus broader goals for schooling (yet)—at this level of generality “competencies” can include appreciation of social diversity, identification with the nation-state, artistic creativity, etc.

completion of basic schooling (Figure 1). There is little basis for working towards universal primary schooling if students emerge from the schooling cycle without an adequate education.

**Figure 1: Key empirical question: Are the actual learning profiles (gain in competency per year of schooling) steep enough that all students who complete the MDG (horizontal axis) really reach the MLG (vertical axis)?**



There is mounting evidence that learning profiles are *not* steep enough and that learning achievement of students in school--even in traditionally measured areas of basic skills such as reading and mathematics (much less conceptual mastery)--is strikingly low.<sup>11</sup>

- A baseline survey of 3<sup>rd</sup> to 5<sup>th</sup> graders in five districts of Andhra Pradesh, a middle performing Indian state, found that only 12 percent of students could do *single digit* subtraction and that 46 percent could not, when shown a picture of six balls and three kites, answer how many kites were in the picture.

<sup>11</sup> We use the results of standardized tests to illustrate the low levels of learning achievement and the examples of competencies from reading, mathematics and science. The education process, of course, is supposed to deliver on other dimensions of individual development such as creativity or other non-cognitive skills. Our discussion of “competencies” is meant to be general—by citing examples of poor performance on arithmetic that arithmetic we do not mean to imply that these should be the most important goal of the schooling system. We do believe, however, that they are one important part of the goals (see discussion in Benavot and Amadio 2004).

- A recent survey of learning in India found that of students in government schools in grades 6-8, who are students who have completed the lower primary cycle and hence met the MDG, 31 percent could not read a simple story, 29 percent could not do two digit subtraction—both of which should have been mastered by grade 2 in the Indian curriculum.
- In Pakistan, tests of grade 3 children found that only 50 percent could answer multiplication questions like “ $4 \times 32$ ” and only 69 percent could not successfully add a word to complete a sentence.<sup>12</sup>
- A recent study in Peru found that “...as few as 25 or 30 percent of the children in first grade, and only about 50 percent of the children in second grade, could read at all” (Crouch 2006).
- In Indonesia, where primary completion is nearly universal 47% of 15-19 year olds could not answer the question “ $56/84 = \dots$ ” correctly.
- In Ghana, a household survey administered eight mathematics questions--where mastery of one digit arithmetic would have been sufficient to answer half the questions and two digit arithmetic to answer all correctly (e.g.  $1+2 =$  ,  $5-2 =$ ,  $2 \times 3 =$ ,  $10/5 =$ ,  $24+17 =$ ,  $33-19 =$ ,  $17 \times 3 =$ ,  $41/7 =$ ) found that only a quarter of 15-19 year olds could answer more than half of these very simple questions.
- In South Africa, 63 percent answered less than half of a set of “real-life” mathematics questions correctly.<sup>13</sup>

A review of the cumulative results of internationally comparable examinations reveals that students in similar grades in rich and poor countries are far apart in learning achievement.

---

<sup>12</sup> Das, Pande, Zajonc (forthcoming).

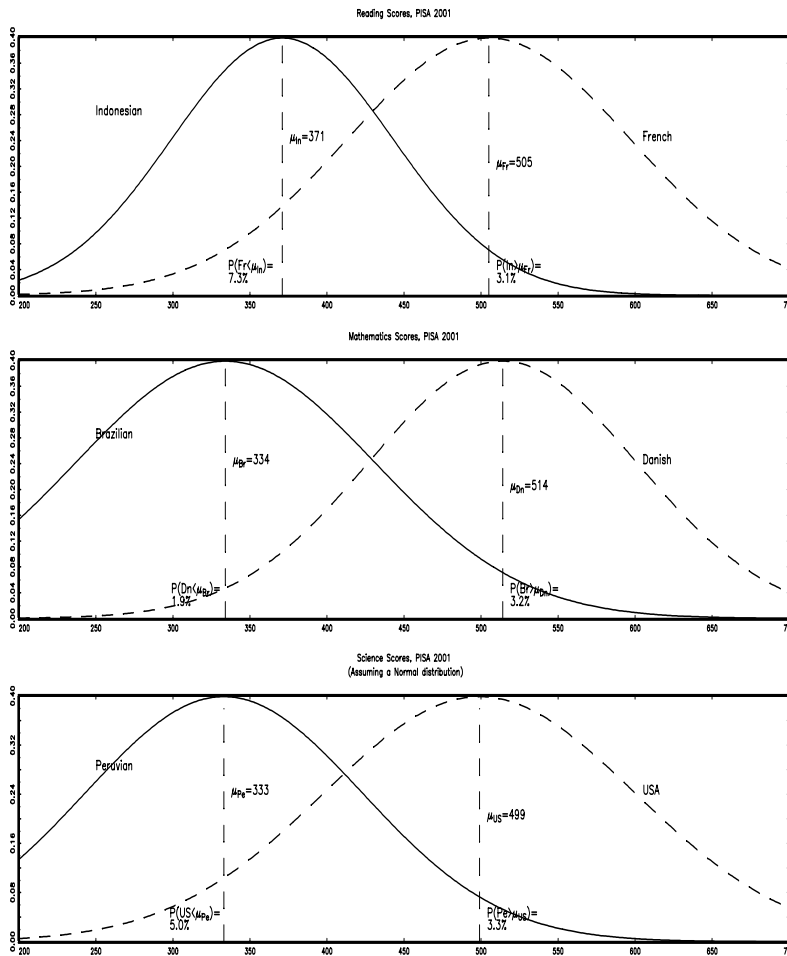
<sup>13</sup> For South Africa, questions were of the following nature: “A shop has 126 litres of milk. 87 litres are sold. How many litres remain?”

For example, Figure 2 shows the distribution of test scores from the Programme for International Student Assessment (PISA) of 2000 for three relatively better off poor countries compared to selected OECD countries.<sup>14</sup> The *average* reading ability of Indonesian students was equivalent to that of the lowest 7 percent of French students. The average mathematics score among students in Brazil was equal to the lowest scoring 2 percent of Danish students. The average science score among students in Peru was equivalent to that of the lowest scoring 5 percent of US students.

---

<sup>14</sup> These curves are simulated on the basis of the mean and the standard deviation for each country.

**Figure 2: Distribution of test scores on the PISA 2000 assessment in reading, mathematics, and science among students from three developing and 3 comparison countries.**



Source: Pritchett (2004)

There is evidence from a variety of countries that even youth that complete basic schooling are leaving school under-equipped to function successfully in a modern (much less global) world. But the fundamental question “are youth getting an adequate education?” cannot be answered: no one knows. No one knows because regular and reliable information on the learning achievement of those children in school is scarce, and information on the competencies of a cohort (both in and out of school) is almost non-existent.



In their measurement of performance international and national policy makers there is a general trend to shift from inputs to outputs and from outputs to outcomes. This is happening in many sectors, including education. The Education for All (EFA) initiative, as defined for example in the goals adopted in 2000 at the World Education Forum in Dakar, Senegal, attempts to explicitly integrate a quality dimension as an objective. The EFA Global Monitoring Report 2005 has quality as its central theme (UNESCO 2004). But all too often an input based approach to school quality is adopted in which measures of inputs thought to be associated with quality are measured (e.g. class size, infrastructure adequacy, teacher qualifications, etc.) but with little or no emphasis on actual student learning. At worst, quality is regarded as an add-on to access to schooling, which inverts the relationship that access to schooling is merely a means to learning.

If the MDG and MLG were “roughly” the same then perhaps the additional effort of defining and creating a new global MLG and even the relatively small effort of countries to define competencies and testing cohorts (not students)—at the very least on a representative sample basis—is not worthwhile. In this paper we examine learning profiles among cohorts of young people in seven developing countries: Brazil, Indonesia, Mexico, Thailand, Tunisia, Turkey and Uruguay.<sup>15</sup> Nearly all have already achieved, or are on track to meet, the MDG target of universal primary completion. We derive two clear results:

- MDG and MLG are *not* closely related. Many students complete their schooling well short of minimal competencies. Achieving the MDG on schooling will leave countries far short of desirable educational goals. Our results show that the majority of youth do not reach a plausible minimal competency level in mathematics, reading and science.

Moreover, the vast majority are nowhere near a global standard of adequate competence.

---

<sup>15</sup> We have not yet been able to identify a household survey data source from which we would be able to derive the highest grade completed among 15 year olds for Tunisia. Because of this we cannot report simulation results for that country, but plan to do so in a subsequent version of this paper.

This underscores the urgent need to move beyond the narrowly defined measures that the MDG currently encourages, to a more relevant outcome: actual acquired competencies.

- Moving from gauging the performance of an educational system from *students* to *cohorts* can make a large difference, and the measured performance of a cohort gives a better picture of educational system performance than quantity based measures (which ignore learning) or exclusively testing the learning of students, which ignores access. We argue that adjusting for the fact that current assessment practice relies on testing only students—rather than youth both in and out of school—is an important step towards building a monitoring system capable of tracking progress towards a Millennium Learning Goal.

The paper is organized as follows: Section 1 briefly describes the state of internationally comparable assessments, and the PISA 2003 data we use in our analysis. Section 2 explains our method for accommodating the fact that only in current students are tested. Section 3 presents our results and alternative estimates to test their robustness and ensure that our results are not driven by selection. Section 4 concludes.

## **1. Assessing student learning and skills**

### **1.1 Internationally comparable student assessment systems**

There are currently several international programs that assess student skills or learning achievement in ways that are comparable across countries. Each has a different background, philosophy, target population, and the different systems cover substantially different countries. The International Association for the Evaluation of Educational Achievement (IEA) has been conducting internationally comparable assessments since 1958. The two most prominent assessments administered by the IEA are the Trends in Mathematics and Science Study (TIMSS)

and the Progress in International Reading Literacy Study (PIRLS). TIMSS and PIRLS are primarily driven by the content of the curricula in the various participating countries, as curricula are used to derive the test items. The Organization for Economic Cooperation and Development (OECD) runs PISA which assesses knowledge and skills needed in adult life, not mastery of the curriculum narrowly defined. It is concerned with the “capacity of students to apply knowledge and skills in key subject areas and to analyze, reason and communicate effectively as they pose, solve and interpret problems in a variety of situations” (OECD 2004). The Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) has run two coordinated rounds of multi-country assessments of students in grade 6, which is TIMSS-like and geared to assessing mastery of the curriculum. The *Programme d’analyse des systèmes éducatifs de la CONFEMEN* (PASEC) has run assessments in nine francophone countries, mainly in West Africa (plus Madagascar), with one-per-country at some point between 1995 and 2001. The *Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación* (LLECE) ran a multi-country assessment of learning outcomes in 13 Latin American countries in 1997 and 1999 geared towards mastery of the curriculum covering students in Grades 3 and 4, the middle of the primary school cycle.

While these assessment systems have all contributed substantially to the knowledge base about student learning in developing countries, they share several shortcomings. First, there is low country coverage—particularly of the poorest countries. SACMEQ and LLECE are considered to be high-quality and regionally relevant programs, but they cover only a fraction of countries in each of their regions (14 out of over 45 for SACMEQ; 13 out of over 25 for LLECE). The coverage of the TIMSS Grade 8 assessment has increased from 15 developing and transition countries in 1995 to 48 such countries in 2003. But coverage remains very low in Sub-

Saharan Africa and East Asia and the Pacific, and there is no assessment in South Asia. Second, it is still early days for most of the assessments, so with the exception of high-income and transition countries, most participating countries have only one data point and therefore limited trend data. Third, these assessment systems have inadequate coverage of various points in the school system. SACMEQ, for instance, is the only internationally comparable program that is focused on monitoring learning outcomes at the close of the primary school cycle—the main current need in the context of Education for All (EFA).

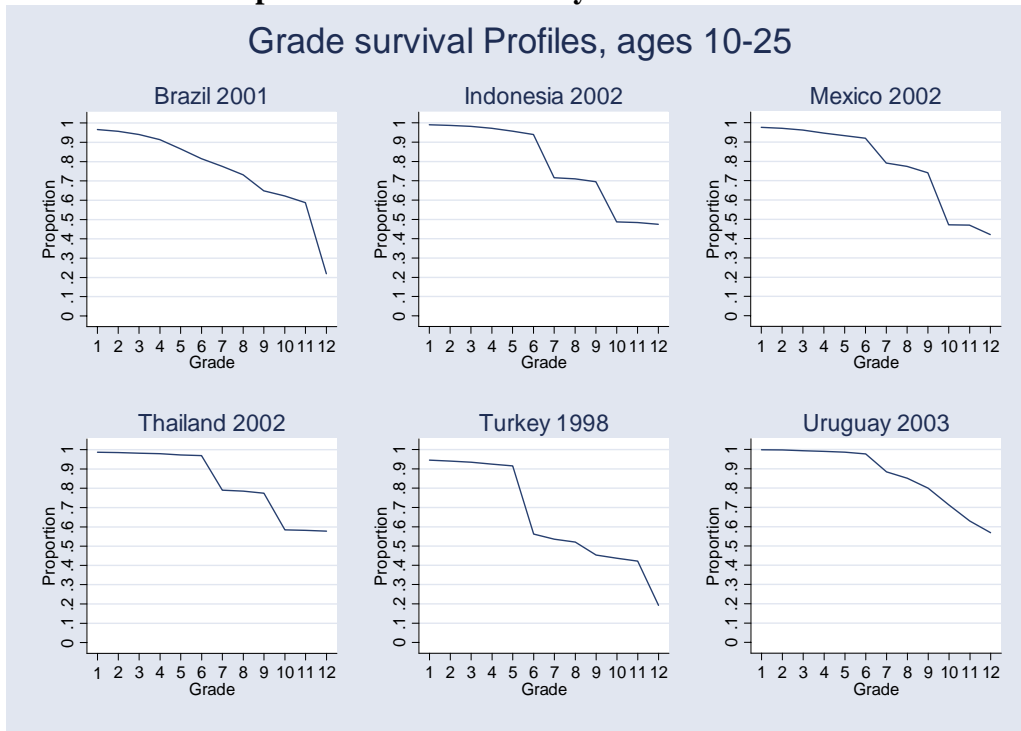
### **1.2 Cohort versus student testing in assessing education system performance**

The main limitation of all existing internationally comparable assessment systems is that they focus exclusively on students who are currently enrolled and attending school. In a country like South Africa, where nearly all pupils continue through to Grade 6, testing students in this grade is not likely to be misleading. But in Malawi where only about 70 percent of pupils make it to Grade 6, many having dropped out in the earlier grades, testing only 6<sup>th</sup> graders means that improvements in learning achieved at lower grades are not being monitored, and that shortfalls in learning and the mastery of skills due to early dropout are not being captured.

The rate at which students drop out of school can vary substantially across countries (Figure 3). In this selected group of countries the grade survival profiles range from a steady pace of dropout across the basic cycle in Brazil, to a flow of dropout after primary completion in Uruguay, to sharper shortfalls across transitions between cycles in most of the other countries (e.g. in Turkey where the basic to secondary transition is very sharp). By Grade 8 for example, the grade in which the TIMSS assessment is made, only about 70% of Brazilian and Indonesian children are still in school, about 80% of children in Mexico, Thailand, and Uruguay. In Turkey (albeit based on somewhat older data) only about 50% of children were still in school by Grade

8.

**Figure 3: The rate of dropout varies substantially across countries**



Source: Authors' analysis of household survey data. Figures show Kaplan-Meier survival curves that adjust for incomplete schooling observations.

Excluding these out-of-school children in assessments of learning and skills can potentially be misleading. Consider, for example, countries where achievement tests were administered in the context of household surveys (Table 1). Discrepancies between testing children currently in school and those in and out of school will depend on the share of children who stay in school, the profile of which grade they drop out in, and learning both in and out of schools. Discrepancies can be relatively small: for example in Ghana or the Cape area in South Africa where including all children lowers the correct response rate by between 3 and 5 percentage points. But they can be quite large: including all respondents in Indonesia, as opposed to just those in school, lowers the average correct response rate by about 10 percentage points.

**Table 1: Household survey results: Scores of Respondents Ages 15-19**

Survey	Year	Respondents In School			All Respondents		
		Maths	Language	Cognitive	Maths	Language	Cognitive
<b>Ghana Living Standards Survey<sup>1</sup></b>	<b>2003</b>	36.1	52.5	62.9	33.5	47.5	60.7
		<i>16.7</i>	<i>23</i>	<i>21.2</i>	<i>19.2</i>	<i>26.7</i>	<i>22.2</i>
		Sample Size	353	354	349	536	536
<b>Cape Area Panel Study<sup>2</sup></b>	<b>2002</b>	50.1	83		46.9	81.1	
		<i>26.3</i>	<i>15</i>		<i>26.5</i>	<i>16.4</i>	
		Sample Size	2,131	2,131		2,820	2,820
<b>Indonesian Family Life Survey<sup>3</sup></b>	<b>2000</b>	57.6		78.4	45		70.4
		<i>31.4</i>		<i>22.8</i>	<i>32.5</i>		<i>27.7</i>
		Sample Size	1,900		1,900	4,048	
<b>Mexico Family Life Survey<sup>4</sup></b>	<b>2000</b>			62.5			56.4
				<i>21.7</i>			<i>23.3</i>
		Sample Size			1,412		
<b>Matlab Health and Socioeconomic Survey<sup>5</sup></b>	<b>1996</b>	84.2			82.5		
		<i>15.1</i>			<i>15.3</i>		
		Sample Size	766			1,079	

Notes: Scores are presented as percent answered correctly. Standard Deviations are in italics; 1 Number of Maths questions = 44, Number of Language Questions = 37, Number of Cognitive Questions = 36; 2 Number of Maths questions = 23, Number of Language Questions = 22; 3 Number of Maths questions = 5, Number of Cognitive Questions = 8; 4 Number of Cognitive Questions = 12; 5 Number of Maths questions = 8;

### 1.3 Conceptual foundations of an MLG

An MLG needs to be based on the notion of *cohort* based assessment of the level of all children of a given age. Consider Figure 4, which returns to the basic notion of a learning goal (on the vertical axis) and a pure completion goal (on the horizontal axis). The example considers four different combinations of schooling attainment and learning profiles:

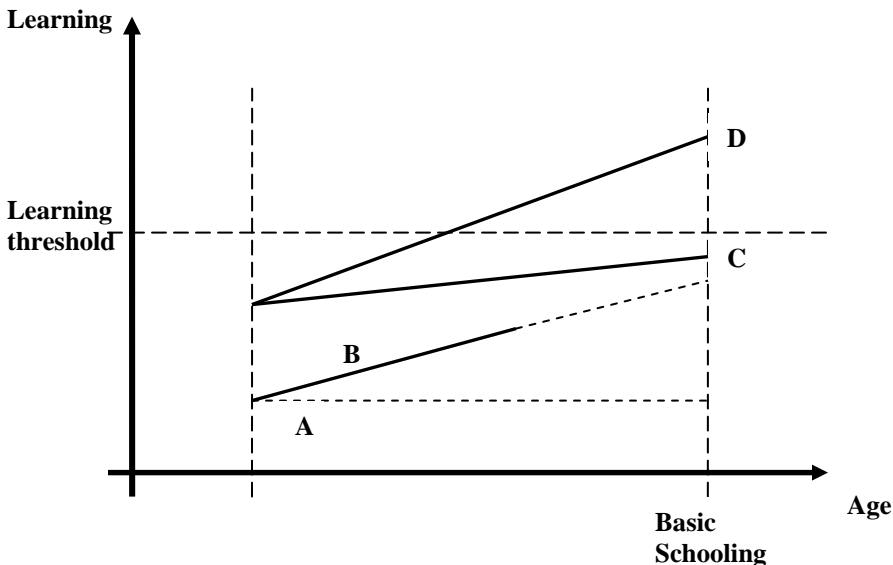
- Student A never enrolls in school and hence is assumed to have a flat learning profile.
- Student B enrolls but drops out before completing a cycle of basic schooling;
- Student C, enrolls and completes a basic cycle of schooling but does not cross a basic learning threshold because the learning profile is too shallow.

- Student D not only completes basic schooling but emerges from the cycle with a level of learning above the minimum threshold.

A literal interpretation of Millennium Development Goal—which is all that is consistently measured and reported in international fora—has allowed energy to be focused largely on getting students such as A to enroll and students such as B to complete the basic cycle. But even if B completes the cycle of basic schooling at the hypothetically illustrated learning profile shown, she would not actually be equipped with the basic competencies necessary to thrive and progress in a modern economy. Similarly, even if student A (who like B illustrates students entering school with low learning readiness) were to enroll and complete the primary cycle he, like B, would also not reach a minimally adequate learning goal.

Moreover, suppose resources were devoted to improving the learning performance of child C and her learning achievement were raised to the level of child D. This gain plays no role in monitoring progress to the MDG. Whether child C completes primary with or without reaching any learning goal (in any subject or any subset of competencies) by the MDG count she is a success. While there is undoubtedly some truth to the claim that the MDGs, and the goals of Education for All more generally, incorporate quality (even if only because, as it is sometimes put: “the MDGs can only be achieved if quality is high enough to attract students”) there has been no systematic attempt to measure and monitor learning progress across a broad range of poor countries—and “what gets measured gets done.”

**Figure 4: Pathways to achieving (or not achieving) a minimally adequate learning threshold.**



An MLG based on cohort based testing can change what is measured to be more consistent with true educational goals. In assessing educational policy the literal application of an MDG framework will be conceptually wrong—and potentially empirically wrong—in cases where analysis based on the cohort distribution of learning achievement will get the right answer<sup>16</sup>. Three examples illustrate the point.

<sup>16</sup> There are, perhaps, cases in which pure attendance may in fact have some benefits. For example, one puzzle is that even in countries in which educational quality appears to be very low there is an impact of maternal education on fertility behavior and of maternal education on child mortality. Perhaps there are some pure socialization gains as just the fact of attending school changes attitudes or behavior even if the student fails to master even basic literacy. However, our view is that this is not a dominant consideration for three reasons. First, most countries are nowhere near on the frontier of effectiveness of schooling and there are large gains in learning achievement possible without trade-offs for access—if these learning goals were pursued. Second, there are also potential negative externalities of schooling without learning as children become alienated and disaffected. Third, many studies suggest that nearly all of the observable gains to wages and to non-economic outcomes from schooling are in fact due to learning and not merely attendance.



First, often low child learning achievement has deep roots, including low school readiness and low initial cognitive ability due to inadequate early childhood inputs (e.g. nutrition, stimulation, etc.). Suppose one were considering an ECD program that raised school readiness so that children who would have completed primary schooling but with low learning now enter schooling with higher learning readiness and would, because of the ECD program, complete primary schooling with (potentially much) higher levels of the valued competencies. An MDG puts zero value on these gains (since the children complete in both cases) where an MLG values the gains in competencies. The rejoinder that some children with the ECD intervention will also complete more schooling and so the MDG does value ECD misses, and therefore makes, the point. The point is that an MDG calculus values ECD *only* as it affects the *quantity* of schooling whereas the true gain to ECD by any reasonable measure of educational progress must include all of the learning gains<sup>17</sup>.

Second, many countries are considering schemes that pay parents to put and keep children in school. Suppose that parents and children have decided to withdraw from school because the child is not making any learning progress. A sufficiently large inducement (e.g. school feeding, conditional cash transfer, scholarship) could induce the parents to force their children to attend in spite of this. Even the worst case scenario of spending government money to induce families to send their children to dysfunctional schools would count as progress towards the MDG—but not towards an MLG. More realistically, suppose four children are in the same school with a shallow learning profile and three of them finish primary school but the

---

<sup>17</sup> While these examples are hypothetical, the trade-offs are real. For instance, many countries have both early child nutrition programs (which potentially have strong effects on size, malnutrition, and cognitive development) and school feeding programs (which act as an inducement to enroll in school but are unlikely to affect nutrition or cognitive status in critical ways). An MDG judging solely on school enrollment will be biased towards school feeding which has attendance and enrollment effects because it cannot properly judge the impact on total student achievement. This list does not pretend to be comprehensive. It focuses on several school-related pathways through which learning achievement may suffer.

fourth child drops out. Suppose that for equivalent cost one could either steepen the learning profile for all four children or induce the fourth child to stay in school. By the MDG standard only the latter policy has *any* gain. Assessed by an MLG the gains to all children (including the learning gain to the fourth child—even if she does not complete the primary cycle) count as gains, as they should.

Third, whenever discussions of educational quality or learning achievement are raised the objection is raised that a focus on “quality” does not properly value access. But this is only true if “quality” measures of learning (vertical axis) are based exclusively on student based testing. In this case one does not have the total distribution of learning achievement of a cohort as the basis for a decision. So, many legitimately worry that a focus on “quality” would perpetuate excessive attention to education for the elite while ignoring the fundamental equity questions. That is, imagine that only students C and D (of Figure 4) actually complete the primary cycle and the “quality” of schooling is judged based on student-only tests of those in Grade 6. Then if student B extends her schooling from Grade 5 to Grade 6 then the measures of access go up but measures of “quality”—average test scores of those tested—would go down. But the distribution of measured learning achievement on a cohort basis would go up (as student B learned more in moving from Grade 5 to Grade 6). Bringing more children into the system may dramatically *raise* cohort learning achievement even while *lowering* observed test scores of those children in schooling.

With measures of the *complete distribution of cohort competencies* the issue of “access” and “quality” is artificial. Since the goal of expanding access is to increase cohort competence the gains of expanded enrollment are represented (which student-only tests miss) and the gains are larger the more the learning while in school (which access measures miss). Assessing the

competencies of both youth in and out of school allows a move away from the debates about quantity versus quality to discussions about policy priorities that improve the overall distribution of competencies. One could have goals for the education system based on the fraction crossing some minimal threshold, the variance, the average, the top end<sup>18</sup>. The debate is not about the relative priorities on “access” versus “quality” but about the relative priorities of raising the low end of the competencies distribution (as would be facilitated by “access” actions that increased enrollments) or middle (steeper learning profiles of those already enrolled) or perhaps at the top end (by more ability- or achievement- versus affordability-based progression to higher tiers of education).

There is the obvious analogy with the distribution of income, in two ways. First, imagine that economists only measured the income of those with wage employment and did not measure incomes of the self-employed (such as peasants). Then a debate about the trade-off between “numbers of jobs” and “wages of those with jobs”—but this false dichotomy would be driven by the artifact of not measuring the complete distribution of income. Second, once there is an estimate of the complete distribution of income across households then one can set various goals or policy objectives based on that distribution of income—goals about poverty, goals about the average level, goals about inequality. In choosing these goals there is a legitimate debate about how gains at various parts of the distribution of income contribute to social objectives: having the complete distribution of income does not imply a focus on the average.

Similarly, with the complete distribution of competencies of a cohort does not mean that only the performance at the top end matters, or the average performance. Rather, having the

---

<sup>18</sup> There is the obvious analogy with the distribution of income, in two ways. First, imagine that economists only measured the income of those with wage employment and did not measure incomes of the self-employed (such as peasants). Then one might imagine there could be a debate about the trade-off between “numbers of jobs” and “wages of those with jobs”—but this would be driven by the artifact of not measuring everyone’s income.

complete distribution turns debates about incommensurables (“access” versus “learning”) into useful discussions of how policy instruments affect the distribution of learning and which learning gains are the priority social objectives.

## **2. Methodology of estimating achievement of an MLG**

In order to illustrate how an MLG might work, we now turn to an illustrative application. However, we do not have cohort based tests of learning achievement suitable for measuring an MLG and will therefore have to use the existing data to estimate, as best we can, what an actual cohort based MLG measurement would produce.

### **2.1 Defining a Millennium Learning Goal**

For the remainder of this paper we will focus our empirical analysis on data from the 2003 round of PISA. In particular, we focus on a selected group of the developing (but not formerly Eastern bloc) countries covered in the assessment exercise (Brazil, Indonesia, Mexico, Thailand, Tunisia, Turkey and Uruguay), as well as several relatively wealthy countries (Greece, Japan, Korea and the USA). These developing countries have either achieved the MDG of universal primary completion or are on track (or close) to achieving the goal by 2015 (see Annex Table 1). Thus our sample can be used to examine how meeting the MDG relates to achieving a possible range of learning goals that could be embodied in a range of MLGs.

PISA covers students who are between ages 15 years 3 months and 16 years 2 months at the time of the assessment, regardless of the grade or type of institution in which they are enrolled and regardless of whether they are in full-time or part-time education.<sup>19</sup> The number of school grades in which these students are enrolled varies depending on national policies on school entry and promotion. In our sample the students are typically in grades 7 through 12.

---

<sup>19</sup> However, if students of this age group were enrolled in primary school they were not included in the study.

Each country sample is typically made up of more than 4,300 students: the smallest sample is for Brazil (4,367) and the largest for Mexico (29,826).

We choose to study the PISA achievement scores for two reasons. First, they allow us to analyze the learning increment across grades since the same test was administered to students in different grades. This is crucial to being able to impute an estimate of cohort achievement. Second, the assessment is not primarily linked to mastery of the curriculum, but to mastery of skills for work and life. If the objective of a Millennium Learning Goal is to monitor progress in preparation for work and life, then the PISA-like assessment is a more appropriate measure (although both types of examinations—curriculum referenced and skills/competencies may continue to be useful for different purposes) and, as we make clear below an MLG does not imply all countries must use the same standards or identical test instruments.

PISA reports levels of competency for mathematics, reading, and science which range from levels 1 (lowest) to 6 (highest). These levels of competency are defined as follows:

- At competence level 1, students can answer questions involving familiar contexts where all relevant information is present and the questions are clearly defined. They are able to identify information and, carry out routine procedures according to direct instructions in explicit situations. They can perform actions that are obvious and follow immediately from the given stimuli. An illustrative level 1 competence question is the following:<sup>20</sup>

---

<sup>20</sup> Available online at [http://www.pisa.oecd.org/document/38/0,2340,en\\_32252351\\_32236173\\_34993126\\_1\\_1\\_1\\_1,00.html](http://www.pisa.oecd.org/document/38/0,2340,en_32252351_32236173_34993126_1_1_1_1,00.html)

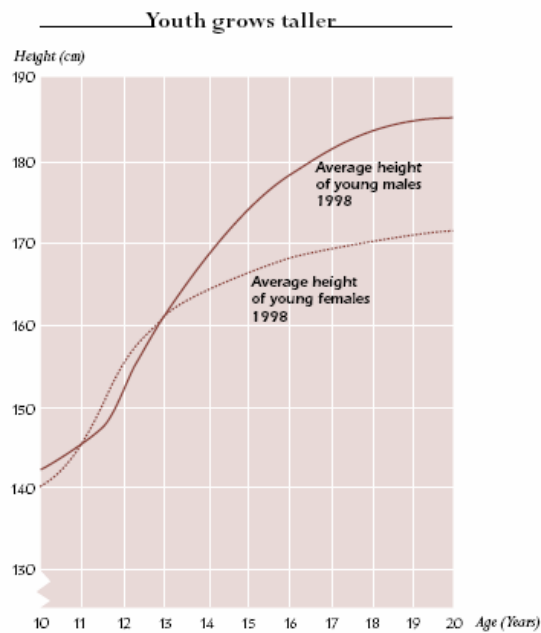
**Illustrative level 1 competence in mathematics question:** *The following table shows the recommended Zedland shoe sizes corresponding to various foot lengths.*

From (in mm)	To (in mm)	Shoe Size
107	115	18
116	122	19
123	128	20
129	134	21
135	139	22
140	146	23
147	152	24
153	159	25
160	166	26
167	172	27
173	179	28
180	186	29
187	192	30
193	199	31
200	206	32
207	212	33
213	219	34
220	226	35

*Marina's feet are 163 mm long. Use the table to determine which Zedland shoe size Marina should try on.*

At competence level 3, students can execute clearly described procedures, including those that require sequential decisions. They can select and apply simple problem-solving strategies. Students at this level can interpret and use representation based on different information sources and reason directly from them. They can develop short communications reporting their interpretations, results and reasoning. An illustrative level 3 competence question presented students with the following figure and question:

**Illustrative level 3 competence in mathematics question:** *In 1998 the average height of both young males and young females in the Netherlands is represented in this graph*



*According to this graph, on average, during which period in their life are females taller than males of the same age?*

We select level 1 competence in reading, mathematics, and science as the lower bound for learning achievement—henceforth referred to as Millennium Learning Goal-Low (or MLG<sub>L</sub>).<sup>21</sup> We do this for several reasons. First, pragmatically, this is the lowest level of competence in the PISA studies and it was meant to define the lowest level that could actually be described as having acquired cognitive skills and conceptual mastery rather than merely mechanical or rote performance. Second, less than 5 percent of OECD students score below level 1. So, for instance, if a worker scoring level 1 or below were to move to an OECD economy they would have a lower competency than 1 in 20 workers. Third, a review of curricular standards in selected developing countries suggests that the curricular mastery of

<sup>21</sup> The PISA scores are normalized to have a mean of 500 across OECD countries. A level 1 competence corresponds to a score of roughly 350 while level 3 competence is close to the OECD mean. Since PISA 2003 did not define these levels for its science assessment we have assigned cutoffs of 350 and 500 for levels 1 and 3 respectively. These are conservative estimates since PISA refers to scores of 400 and 690 as low and top end of the science scale. See OECD (2004), p.292.

primary education is assumed to produce at least level 1 competencies, and many are much more ambitious.<sup>22</sup> In the implementation of an MLG approach each country will of course be free to adopt its own definition of a national MLG (in addition to global standards), but we expect no country would wish to choose a lower goal as a target<sup>23</sup>. Producing students who can demonstrate level 1 competencies in reading, mathematics and science is a reasonable minimal target for a functioning school system.

As an upper bound for learning achievement we chose the level of 500—henceforth Millennium Learning Goal-High (or MLG<sub>H</sub>). We select this as it corresponds roughly to the OECD mean score, and is a realistic upper bound target for average skill mastery. Some might argue this standard is “too high” as it expects poor countries to achieve the levels of OECD countries—even though this might not be needed given the country circumstances. But we retain it for several reasons. First, in the context of a complex and globalized economy even OECD countries are worried that their educational systems are failing to produce well prepared graduates. In many rich-country labor markets those with only average skills have been losing substantial ground in employability as ever higher level of skills are required.<sup>24</sup> Second, this is

---

<sup>22</sup> At the lowest performance level of Brazil’s National Basic Education Evaluation System (SAEB), students are expected to be able to undertake “Object identification and determination, data interpretation through bar diagrams analysis, and identification of simple geometric figures (Guimarães de Castro 2001). The goals of reading instruction at the primary level in Turkey for instance are to “have students gain the ability to accurately understand what they observe, listen to, and read.” They also seek to “have students gain accurate explanatory skills and habits in spoken and written form, based on what they see and observe, listen to, read, examine, think, and plan (Atliođlu 2002). In Tunisia, for instance, students must score an average of 50 percent or better on regional examinations at the end of the sixth grade to progress to the lower-secondary cycle. Students are tested in Arabic writing and reading comprehension, French writing, reading and dictation, mathematics, introductory science, Islamic studies, history, geography and civic education.

<sup>23</sup> A perhaps useful analogy is the World Bank’s use of poverty lines. There are international lower bound standards to examine global progress, less clarity on an upper bound for poverty (see Pritchett 2006) but each country also uses its own poverty line based on country specific calculations.

<sup>24</sup> Murnane and Levy (1996) for instance argue that to be employable in the USA at wages that can support a family a worker needs: The ability to read and do math at the ninth-grade level (roughly the level 3 competencies); and the ability to solve semi-structured problems where hypotheses must be formed and tested, the ability to work in groups with persons of differing backgrounds, the ability to communicate effectively, both orally and in writing; and the ability to use computers to carry out simple tasks like word processing.



the *median* of OECD 15 year olds—many of whom expect to ultimately have much more education—so this allows for different overall targets for the education system, targets that are linked to ultimate level completed. Our MLG-High threshold is therefore not suggesting that poor countries produce students at the OECD median of the labor force at completed schooling, but rather that the target for OECD competencies already achieved by age 15 should be comparable for global progress. Third, if the TIMSS tests are in any way an indication (and there are many issues with comparability) then many developing countries—including China—are in fact producing students who mostly meet or exceed this standard. As we see below, only roughly a quarter of a recent Korean cohort of students do not reach this level. So this is not impossible for countries to achieve even with limited resources.

Again, in practice the goal is to create nationally and internationally accepted standards and assess the progress of entire cohorts towards those standards<sup>25</sup>. Once we have the entire distribution of learning achievement for a cohort measured over time it is relatively straightforward to track progress against an international lower bound, an international upper bound, and each country can track national goals.

## **2.2 Estimating performance where it is missing**

In order to calculate the fraction of students below a given score we need to estimate the distribution of scores of the 15 year olds who did *not* take the test. We illustrate the approach by describing the simplest case: estimating just the average level of performance. We then elaborate on the approach, and estimate the entire distributions of scores for each grade attained, which ultimately to determine the fractions of the cohort that lie below the two MLG thresholds.

---

<sup>25</sup> This needn't imply that every country adopt exactly the same test instrument, as countries could introduce additional material on subjects of national importance or additional assessments to measure "higher order" skills (such as creativity). As long as a sufficient core of comparable items were retained it should be possible to have variation in national assessment instruments while maintaining international comparability.

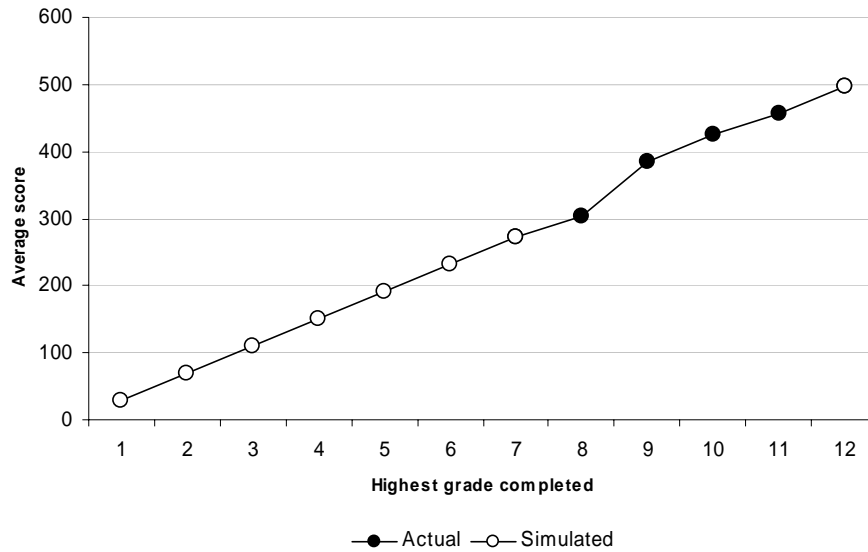
*Estimating the average achievement by highest grade attained of non-test takers*

We know what 15 year old students enrolled in grades 7-12 scored on the PISA subject tests. What we want to find out is what youth who were not in school *would have scored* had they taken the PISA tests. Consider the case where we only want to estimate the average level of learning. We exploit the fact that we observe students in multiple grades and calculate the grade to grade increments in learning for each country (we return to questions of selection effects in the next section). We then use the median of these increments to interpolate back to the lowest level of grade attainment—including never enrolled. For example in Brazil (as illustrated in Figure 5) we observe test-takers in grades 7 through 11. The median performance increment in these grades is 41 points per year (which is roughly 43% of a standard deviation across individuals) so the average score increased from 272 of those who take the test in grade 7 to 457 for those in grade 11. We then apply this increment recursively to interpolate “back” to Grade 0 (never enrolled). In addition, we apply the increment to interpolate “forward” to Grade 12.<sup>26</sup>

---

<sup>26</sup> Linear interpolation is a simplification. As discussed below, the results are robust to alternative approaches to interpolation.

**Figure 5: Actual math performance in Brazil and simulation based on linear interpolation.**



Source: Authors' analysis of PISA data

Once we have the actual and estimated average score by highest grade attained, we then estimate the percent of 15 year olds who have attained each grade, as derived from an analysis of household survey data.<sup>27</sup> Multiplying the average score (S) in each grade by the proportion of the cohort at each highest grade attained (Prp(C=G)) and summing over all grades (0 to G) yields the overall average performance (M) as described, that is:

$$M = E(S) = \sum_{G=0, \max} \text{Prp}(C=G) * E(S | C=G)$$

The distribution of grade attainment among 15 year olds is the weight used to estimate the *cohort* average PISA score.<sup>28</sup>

*Estimating the entire distributions of test scores for non-test takers*

<sup>27</sup> The surveys we use are: Brazil Pesquisa Nacional por Amostragem de Domicilios 2001 (PNAD 2001), Indonesia National Socioeconomic Survey 2002 (SUSENAS 2002), Mexico Encuesta Nacional de Ingresos y Gastos de Hogares 2002 (ENIGH 2002), Thailand Socioeconomic Survey 2002 (SES 2002), Turkey Demographic and Health Survey 1998 (DHS 1998), Uruguay Encuesta Continua de Hogares 2003 (ECH 2003).

<sup>28</sup> See Annex Table 2 for the numerical details for each country. Note that for consistency we use the percentage with each highest grade attained derived from household surveys, including for the grades covered by PISA.

In order to be able to estimate the fraction of a country's cohort meeting the MLGs, we need to approximate the entire *distribution* of performance. We use a similar approach to the one we use for the mean. The fraction (F) of a given age cohort reaching the MLG threshold equals the sum over each grade (0 to G) of the product of the proportion that has achieved at most that grade ((Prp(C=G)) and the proportion among them who score above the MLG (Prp(S>MLG)), that is:

$$F = \text{Prp}(S > \text{MLG}) = \sum_{G=0, \text{max}} \text{Prp}(C=G) * \text{Prp}(S > \text{MLG} | C=G)$$

The first step is therefore to estimate the fraction above the threshold at each grade. The key to this is the assumption that scores are normally distributed and have a constant coefficient of variation across grades.<sup>29</sup> Using the grade-by-grade means estimated above we calculate the standard deviations associated with those means using the country specific estimate of coefficient of variation. Once we have the mean and the standard deviation at each grade, we calculate the implied proportion below the MLGs by applying standard distribution functions to the normalized distribution.

### **3. Results**

#### **3.1 Main results**

The percentage of test takers below  $\text{MLG}_L$  on the math assessment is as low as 2 percent in Korea and as high as 64 percent in Brazil (Table 2).<sup>30</sup> When the distribution of test scores for the entire cohort of 15 year olds is simulated there is a sharp increase in the percentage of 15 year olds who fall below  $\text{MLG}_L$ . In the starkest case, Turkey, there is a 21 percentage point

---

<sup>29</sup> This assumption is generally borne out by the data.

<sup>30</sup> Slight differences from OECD (2004), p. 354 are due to the fact that we consistently use the distribution of grade attainment as derived from household surveys, even for the grades for which there are PISA data. This results in a slight re-weighting of the PISA results across grades. Using the distribution across grades implicit in the PISA test sample does not qualitatively change our results.

increase in the percentage of the *cohort* scoring below  $MLG_L$  compared to the test takers. Recall that  $MLG_L$  corresponds roughly to level 1, the minimum level of competency. Turkey is “on track” to achieve universal primary completion by 2015, yet two-thirds of 15 year olds are unable to perform at the minimum competence level as identified by PISA.

Among the two countries that are off-track to meet the MDG of universal primary completion, Thailand appears to be imparting better quality to more students than Indonesia: only 34 percent of Thai 15 year olds compared to 68 percent of Indonesians fall below the minimum threshold. Thus for every Thai student below the  $MLG_L$  there are two Indonesian students who fail to meet the threshold.

If one considers the higher Millennium Learning Goal, at level 3 where students “can select and apply simple problem-solving strategies,” fully 91 percent of the Turkish cohort of 15 year olds fails to reach the goal. More than 90 percent of 15 year olds fail to reach the  $MLG_H$  in Brazil, Indonesia, Mexico and Turkey.

An analysis of the scores of female students reveals that they perform worse on the math assessment in all countries, regardless of whether one focuses on test-takers or at the cohort of 15 year old females. The sole exception is Thailand where females do marginally better on  $MLG_L$  (24 percent of female test takers below rather than 26 percent of all test takers) and no worse when we consider  $MLG_H$  (80 percent in either case).<sup>31</sup>

---

<sup>31</sup> Detailed tables that summarize our findings for females are included in the annex.

**Table 2: Percentage Below MLG-Low and MLG-High in Mathematics**

	Test Takers		Cohort		Female Cohort	
	Percent Below	Percent Below	Percent Below	Percent Below	Percent Below	Percent Below
	MLG-Low	MLG-High	MLG-Low	MLG-High	MLG-Low	MLG-High
Brazil	64	93	78	96	78	97
Indonesia	59	97	68	98	69	97
Korea*	2	25	2	25	2	30
Mexico	38	88	50	91	52	92
Thailand	26	80	34	82	30	82
Tunisia						
Turkey	45	84	67	91	74	94
Uruguay	32	77	39	79	40	81
Greece*	17	66	17	66	18	71
Japan*	3	30	3	30	3	30
USA*	9	49	9	49	8	51

\* Enrollment of 15 year olds assumed to be 100 percent in the grades covered by PISA for these countries.

All of the countries analyzed perform better on the reading assessment than on the math assessment (Table 3). The percentage of test takers below the  $MLG_L$  is highest in Brazil at 33 percent, however we estimate that 57 percent of the population of 15 year olds fails to reach this goal. In Indonesia virtually no 15 year olds meet the higher MLG in reading: 97 percent fall below the mark. While Indonesia may be close to achieving universal primary education by 2015, almost none of its students are able to meet a global competency standard.

The results for female test takers are more promising on the reading assessment. Female test takers do better on meeting both  $MLG_L$  and  $MLG_H$  in all countries except Turkey where they perform marginally worse: two percentage points and four percentage points respectively. In the population of 15 year olds as a whole however, the results are less reassuring for Turkey. The proportion of Turkish females in the population of 15 year olds who fall below  $MLG_L$  is fourteen percentage points higher than the overall population.

**Table 3: Percentage Below MLG-Low and MLG-High in Reading**

	Test Takers		Cohort		Female Cohort	
	Percent Below	Percent Below	Percent Below	Percent Below	Percent Below	Percent Below
	MLG-Low	MLG-High	MLG-Low	MLG-High	MLG-Low	MLG-High
Brazil	33	84	57	90	48	88
Indonesia	31	96	45	97	37	96
Korea*	0	24	0	24	0	18
Mexico	24	82	39	86	35	84
Thailand	13	80	19	83	12	77
Tunisia						
Turkey	28	81	50	89	64	93
Uruguay	24	70	31	74	23	69
Greece*	8	54	8	54	4	46
Japan*	5	43	5	43	3	38
USA*	5	43	5	43	2	36

\* Enrollment assumed to be 100 percent in the grades covered by PISA for these countries.

Performance in the science assessment lies between those in mathematics and reading. Over half of 15 year olds in Brazil and Turkey fail to reach  $MLG_L$ , between 25 and 40 percent in Indonesia, Thailand and Uruguay.  $MLG_H$  continues to be a harder hurdle: over 90 percent of 15 year olds fail to reach the goal in Brazil, Indonesia, Mexico, and Uruguay.

The performance among the female cohort is similar to that for mathematics. Females in all countries do worse than the population as a whole except in Thailand where their performance is marginally better.

**Table 4: Percentage Below MLG-Low and MLG-High in Science**

	Test Takers		Cohort		Female Cohort	
	Percent Below MLG-Low	Percent Below MLG-High	Percent Below MLG-Low	Percent Below MLG-High	Percent Below MLG-Low	Percent Below MLG-High
Brazil	43	93	64	96	62	96
Indonesia	28	98	39	98	41	98
Korea*	2	34	2	34	3	38
Mexico	25	89	38	91	40	93
Thailand	17	83	26	85	22	84
Tunisia						
Turkey	33	88	57	93	67	95
Uruguay	24	77	31	80	29	81
Greece*	7	59	7	59	2	32
Japan*	3	32	3	32	6	55
USA*	7	53	7	53	8	62

\* Enrollment assumed to be 100 percent in the grades covered by PISA for these countries.

Table 5 is the summary table of the “base case” estimates of how many 15 year olds are not reaching potential MLG target levels on basic competencies. The enormous gaps between the well-performing and badly performing countries are striking. Only 5 percent of US 15 year olds were not above the MLG-Low in reading—and the bottom 5 percent of US students are not generally considered to be at an acceptable level of functional literacy. In Brazil fully 57 percent did not reach that standard, 50 percent of 15 year olds in Turkey, 45 percent in Indonesia.

The numbers for the OECD median (as a possible MLG-High) are similarly striking: only 4 percent of students in Brazil could do science as well as the typical US student—or the bottom third of Korean or Japanese students. These estimates suggest the *top ten percent* of students in mastery of Mathematics in Brazil, Indonesia, or Mexico do not reach the performance of the OECD median.



**Table 5: Percentage of a Cohort of 15 year olds estimate to be below MLG-Low and MLG-High in Mathematics, Reading and Science (sorted from worst average on MLG-Low to best)**

	Below MLG-Low			Below MLG-High		
	Mathematics	Reading	Science	Mathematics	Reading	Science
Brazil	78	57	64	96	90	96
Turkey	45	50	57	84	89	93
Indonesia	68	45	39	98	97	98
Mexico	50	39	38	91	86	91
Uruguay	39	31	31	79	74	80
Thailand	34	19	26	82	83	85
Greece*	17	8	7	66	54	59
USA*	9	5	7	49	43	53
Japan*	3	5	3	30	43	32
Korea*	2	0	2	25	24	34

\* Enrollment assumed to be 100 percent in the grades covered by PISA for these countries.

If these numbers reveal a deep disconnect between accomplishing a quantity target for years of schooling completed and education actually achieved, it should be kept in mind these countries are likely to provide an optimistic view. These are mostly middle income countries that have (or will likely) reach the MDG. If Turkey, which has begun EU negotiations, can manage only 50 percent achieving higher than a minimal competency in reading, one can only imagine (because there are no comparable results) how awful measured learning might be for poorer countries with weaker institutions.

### 3.2 Robustness and selection effects

To this point we have operated on the assumption that a linear interpolation of the observed scores across grades is a valid approximation of a *learning* profile. In doing so, we have no doubt irked many econometricians who spend their days, and some of their nights, trying to overcome problems such as measurement error and selection bias. This section is meant to assuage this audience. Non-technical readers may wish to skip to the concluding section with

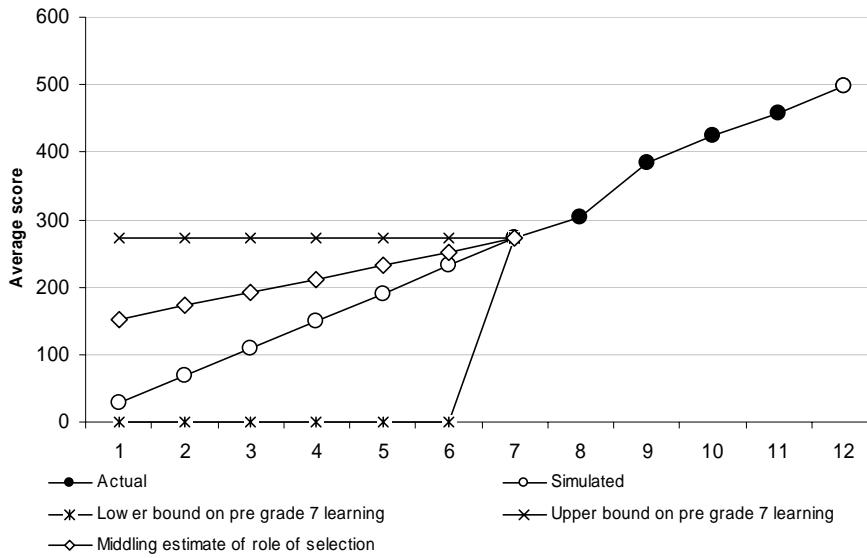
the comfort that this section suggests that the empirical results are not overly sensitive to reasonable assumptions about either of these potential problems.

The problem of measurement error is straightforward: is our estimate of the learning increment adequate, or is it sensitive to the sample or grades it is based on? The problem of selection would arise if those who were tested differ systematically from those that were not. The result would be that we would be estimating a biased estimate of the true learning growth and therefore inferring a biased learning profile. What might the extent of this bias be?

As a first approach, consider the bounds of the extent of the bias. It is possible that our interpolated line is an understatement of learning gains and that therefore the line should be much steeper than we have shown it to be. The “systematic difference” between test takers and non-test takers is that non-test takers (i.e. those who have dropped out) have substantially lower than predicted competencies. An upper bound for this would be if all the gain accrued in the year prior to the one for which we observe data. In other words, in this extreme, the first 6 grades really do not impart anything to the student. Figure 6 illustrates this scenario with a line that runs along the x-axis until Grade 6 and then rise steeply to the observed score at Grade 7.

It is also possible that we have overestimated the year to year learning gain in which case the interpolated line should be much flatter than we have shown it to be. As illustrated in Figure 6, in the extreme the learning profile for these early grades would be a flat line extending from the score for grade 7 (the lowest grade we typically observe) to the y-axis.

**Figure 6: Actual and simulated mean math performance in Brazil**



Source: Authors' analysis of PISA data

We use three alternative approaches to estimating learning gains which address robustness and selection.<sup>32</sup> The first approach addresses mainly the robustness issue: we use different approaches to averaging across the various increments we derive across pairs of grades for each country (median, mean, highest, lowest, using only the pair of grades with the largest number of test-takers) but our results are not sensitive to these different approaches. The second and third approaches address primarily the selection issue, but are also additional robustness checks. The second approach consists of calculating the learning increment using only those students who were in the highest economic status quintile.<sup>33</sup> Because dropout is minimal in this subset of the population, selection should not be an issue—or at least should be a much more minor issue. The third approach is to isolate the exogenous learning gain using the age for grade variation in the data. Recall that PISA examines those between the ages of 15 years and 3 months

<sup>32</sup> Recall that we only calculate an increment if it there two adjacent grades where both had a sample size greater than 50, increasing the stability of our results. We also experimented with estimating the cohort distribution across highest grade completed in the household surveys using both 15 and 16 year olds, but the results were not sensitive to this change.

<sup>33</sup> As measured by an index of consumer durables owned by family members.

and 16 years and 2 months. We divide the sample into two groups – those who are young (below 15.75 years of age) and those who are old (above 15.75 years of age). We then calculate the grade increment between those who are young in one grade and those who are old in the next grade. Thus we isolate the exogenous age-based part of learning. Armed with this learning increment, we then re-run the analysis.

The results from these alternative approaches are reported in Tables 6 and 7 (where they are compared).

Our results are largely unchanged after changing the way we derive the increment. For example the percentage of Brazilian 15 year olds not meeting the minimal level of competency in math is 79 percent using our basic approach (Table 2, repeated in the first column of Table 6 for comparison), compared to 78 percent when the increment is derived from quintile 5 test-takers, and 74 percent when using only the exogenous age for grade increase. This overall consistency in results carries over across countries, across to the reading and science results, and across to the fraction reaching the higher level of competency. One might worry that these results are “too” similar. Recall that we are inferring the tail of a distribution. In many cases, particularly the MLG-high results, that tail is so far from the mean that (relatively) small changes in the mean have little effect on the estimate of the share in the tail.

**Table 6: Percentage of cohort below MLG-Low using different assumptions to estimate the learning profile**

	Math			Reading			Science		
	Basic	Quint- ile 5	Age/ Grade	Basic	Quint- ile 5	Age/ Grade	Basic	Quint- ile 5	Age/ Grade
Brazil	78	78	74	57	59	56	64	65	63
Indonesia	68	69	67	45	47	48	39	43	42
Korea*	2	-	-	0	-	-	2	-	-
Mexico	50	50	49	39	39	39	38	38	37
Thailand	34	34	31	19	19	23	26	26	25
Tunisia									
Turkey	67	65	70	50	46	50	57	48	58
Uruguay	39	39	36	31	32	33	31	31	31
Greece*	17	-	-	8	-	-	7	-	-
Japan*	3	-	-	5	-	-	3	-	-
USA*	9	-	-	5	-	-	7	-	-

\*Enrollment assumed to be 100 percent in the grades covered by PISA for these countries.  
 - Indicates that no modeling was applied and therefore simulation results reported.

**Table 7: Percentage of cohort below MLG-High using different assumptions to estimate the learning profile**

	Math			Reading			Science		
	Basic	Quint- ile 5	Age/ Grade	Basic	Quint- ile 5	Age/ Grade	Basic	Quint- ile 5	Age/ Grade
Brazil	96	96	95	90	90	90	96	96	96
Indonesia	98	98	97	97	97	97	98	98	98
Korea*	25	-	-	24	-	-	38	-	-
Mexico	91	91	90	86	86	86	93	91	91
Thailand	82	82	81	83	83	83	84	85	85
Tunisia									
Turkey	91	91	91	89	89	89	95	93	93
Uruguay	79	79	78	74	74	74	81	80	79
Greece*	66	-	-	54	-	-	32	-	-
Japan*	30	-	-	43	-	-	55	-	-
USA*	49	-	-	43	-	-	62	-	-

\* Enrollment assumed to be 100 percent in the grades covered by PISA for these countries.  
 - Indicates that no modeling was applied and therefore no simulation results reported.

#### **4. Conclusion and Direction Forward**

This paper simply demonstrates that accomplishing the MDG will not leave the youth of developing countries well-prepared for lives as adults in the 21<sup>st</sup> century. While the fact that learning achievement in many developing countries of those children in school is very low has been established in many international comparisons, we argue that a cohort learning approach is conceptually superior. While we illustrate how an MLG might work using a simulation based on existing data, we believe that collecting the relevant cohort-based data would be far superior.

We recommend that international agencies and individual countries should move from MDG target on completion to an MLG target on cohort learning achievement and mastery of competencies. This is a continuation of a natural transition, from MDG to “MDG with quality” to MLG<sup>34</sup>. The MDG approach has already served a very useful purpose in focusing attention of schooling deficits and bolstering the notions of output targets and accountability—but ultimately falls short of capturing the actual goal of education for all. Individual countries can and should adopt the MLG approach on their own—there is no need to wait for the international system to catch up with countries that are able to move ahead.

While the MDG and EFA approaches focused on enrollment and completion have been useful, there are three problems with sticking too long to the existing MDG approach. First, a focus on the quantity of school attendance too often leads to the view that there is an “unmet need” and that there is a simple “solution” which consists of technocratic, logistical actions such as buildings schools, hiring and training teachers, and, simply getting children to report for

---

<sup>34</sup> Simply adding “quality” as an additional concern to the MDG is a useful step, but is inadequate as a coherent framework, especially compared to the MLG which integrates quantity and quality with cohort based measurement, as it does not produce a coherent way of deciding among goals or policies.

school.<sup>35</sup> As a result, to the extent that there is accountability within the system, it involves the easily observable tasks—even if they have little or no relation to the *real* objective of learning. As political scientists have pointed out, the physical expansion of systems or the expansion of expenditures have powerful coalitions (e.g. contractors and providers) while promoting learning, and the incentive structures necessary for it, is a much more difficult task (Grindle 2004).

Second, as the *World Development Report 2004: Making Services Work for Poor People* also highlights, this disconnect between politically expedient observable interventions, and the harder to implement and harder to attribute interventions that may actually have larger impacts on outcomes (World Bank 2003). There is increasing acceptance that a large part of what it will take to increase learning is to establish an appropriate system of accountabilities within the education sector to create a performance orientation around learning outcomes (as discussed more extensively in World Bank 2003). But the MDGs set up a global accountability mechanism that, perhaps inadvertently but inexorably nonetheless, focuses *only* on the quantity of students in classrooms. This invites, if not demands, a bureaucratic accountability for quantity only.

Third, the lack of a clear measurable goal centered around learning, in the face of just such a goal for attendance, has the potential to distort policymaking and ultimately undermine the long-term interests of the countries that are trying to meet such goals. Our analysis indicates that while indicators of the quantity of education have improved, they do not hold up to a more nuanced examination which introduces even a minimalist measure of learning achievement or competencies. Moreover, as many countries meet the MDG target, sticking to the MDG might create the erroneous impression that the education agenda loses priority when all children make it through the last grade of primary school.

---

<sup>35</sup> See the discussion in Pritchett and Woolcock (2004).

There are three steps to be taken for a country, or an international institution, to adopt an MLG.

- Each country (or set of countries) must define a realistic set of competencies as their “low” and “high” targets for learning. While naturally nearly all countries have written curricula, these are often over-ambitious, over complicated, overly broad, and not linked to specific competencies expected from mastery of the curriculum<sup>36</sup>. The fact that in many countries achievement is so far from specified curricular goals suggest they are out of touch with on-the-ground reality of the education system (in part because these goals were often set for an education system that was expected to cater only to an elite). While we have used tests of reading, mathematics, and science from PISA this is only an illustration. Since education is about the socialization of the next generation of citizens, societies should be free to set whatever goals they choose. That said, we believe that nearly all countries would include standards for these subjects: it is hard to see a social consensus developing that excludes functional literacy, command of basic numerical reasoning, and understanding of basic notions of science, as important elements of a universally desirable education.
- Countries would have to agree how to measure the desired competencies from schooling on a regular basis. This is not to say that this testing has to be done for every child—sample based testing can accurately measure overall performance of the system. Moreover, these competency assessments need not replace existing national examination systems, which typically have the entirely different purpose of rationing access to higher levels of schooling, and could be a supplement to rather than replacement of national

---

<sup>36</sup> In the WDR 2004 terminology there is weak “delegation” in the accountability relationships as the goals of school are often contested and hence are left ambiguous, but ambiguous and diffuse goals are interpreted in practice, often in ways inconsistent with social interests.



assessments which monitor mastery of the curriculum among students. In addition, each country will have to decide whether these competency based assessments become “high stakes” for schools or units of the schooling system. This is possible but not a necessary element. Moving to an MLG will require a discussion of the specific competencies that should be promoted and how they should be measured. Fortunately, on this there has been enormous recent progress in the context of existing international assessments.

- While there are a number of internationally comparable tests, all of the international tests assess only students who are in school. Beyond the fact that this makes international comparisons invalid, this also implies that test results are not an adequate measure of progress—the relevant indicator is the competencies of a cohort, whether they are in school or not. Almost no schooling system consistently measures progress towards *outcome* based goals such that the politicians and policy makers with accountability to citizens for their nation’s education system can report on whether they are meeting any universal target for learning.<sup>37</sup>

The question of the past was “can we get all children in school?” while the question now is “are youth emerging from the educational system adequately equipped for their future?”

---

<sup>37</sup> For a recent debate on the issue in Tunisia see Akkari (2005). Also see discussions in Abadzi and others (2005) and Motivans (2005).

## References

The word processed describes informally reproduced works that may not be commonly available through libraries.

Abadzi, Helen, Luis Crouch, Marcela Echegaray, Consuela Pasco, and Jessyca Sampe. 2005. "Monitoring Basic Skills Acquisition Through Rapid Learning Assessments: A Case Study from Peru." *Prospects* 35(2):137–156.

Akkari, Abdeljalil. 2005. "The Tunisian Educational Reform: From Quantity to Quality and the Need for Monitoring and Assessment." *Prospects* 35(1):59–74.

Atliođlu, Yardunar. 2002. *PIRLS 2001 Encyclopedia: A Reference Guide to Reading Education*. Chestnut Hill, M.A.: International Study Center, Lynch School of Education, Boston College.

Benavot, Aaron, and Massimo Amadio. 2004. "A Global Study of Intended Instructional Time and Official School Curricula, 1980-2000.". Processed.

Bruns, Barbara, Alain Mingat, and Ramahatra Rakotomalala. 2003. *Achieving Universal Primary Education by 2015: A Chance for Every Child*. Washington, DC: World Bank.

Crouch, Luis. 2006. "Education Sector: Standards, Accountability, and Support." In David Cotlear, (eds.), *A New Social Contract for Peru: An Agenda for Improving Education, Health Care, and the Social Safety Net*. Washington, D.C.: World Bank.

Das, Jishnu, Rohini Pandea and Tristan Zajonc. Forthcoming. "Learning Levels and Gaps in Pakistan." World Bank Policy Research Working paper Series.

Grindle, Merilee. 2004. *Despite the Odds: The Contentious Politics of Education Reform*. Princeton, NJ: Princeton University Press.

Guimarães de Castro, Maria Helena. 2001. *The National Basic Education Evaluation System - SAEB*. Atlanta, GA: Education Partnerships Implementation Commission (EPIC).

Motivans, Albert. 2005. "Using Education Indicators for Policy: School Life Expectancy." *Prospects* 35(1):109–116.

Murnane, Richard J., and Frank Levy. 1996. *Teaching the New Basic Skills: Principles for Educating Children to Thrive in a Changing Economy*. New York, NY: The Free Press.

OECD. 2004. *Learning for Tomorrow's World: First Results from PISA 2003*. Paris, France: Organization for Economic Co-operation and Development.

Pritchett, Lant. 2004. "Access to Education." In Bjørn Lomborg, (eds.), *Global Crises, Global Solutions*. Cambridge, UK: Cambridge University Press.

Pritchett, Lant, and Michael Woolcock. 2004. "Solutions When *the* Solution is the Problem: Arraying the Disarray in Development." *World Development* 32(2):191–212.

UNESCO, 2004. *Education for All Global Monitoring Report: The Quality Imperative*. Paris: UNESCO Publishing.

World Bank. 2003. *World Development Report 2004: Making Services Work for Poor People*. Washington, DC: World Bank.

-----, 2006. *World Development Report 2007: Development and the Next Generation*. Washington, DC: World Bank.

**Annex Table 1**

<b>Country</b>	<b>Years in Primary Cycle</b>	<b>Universal Primary Completion by 2015</b>
Korea	6	Achieved
Mexico	6	Achieved
Uruguay	6	Achieved
Brazil	8	On Track
Tunisia	6	On Track
Turkey	5	On Track
Indonesia	6	Off Track (but close)
Thailand	6	Off Track (but close)

Source: Bruns, Mingat, and Rakotomalala (2003)

**Annex Table 2 – National Average PISA Score decomposed**

	Grade Attainment Distribution of Cohort of 15 year olds													Mean Interpolated Using Median Increment across Grades													Cohort Average
	0	1	2	3	4	5	6	7	8	9	10	11	12	0	1	2	3	4	5	6	7	8	9	10	11	12	
Brazil	2	2	2	4	7	10	12	14	21	21	4	1	0	0	27	68	109	150	190	231	272	303	384	425	457	498	267
Indonesia	1	0	1	2	2	2	17	5	18	39	10	2	1	124	149	174	200	225	251	276	301	316	348	395	413	353	322
Korea	0	0	0	0	0	0	0	0	0	0	100	0	0	440	451	461	471	481	491	502	512	522	532	542	601	611	542
Mexico	2	0	1	2	2	2	12	5	10	30	33	2	0	0	33	75	117	159	201	243	285	327	369	422	456	552	346
Thailand	1		0	0	0	0	8	2	5	43	32	5	2	0	0	8	72	136	200	264	328	351	395	434	523	587	394
Tunisia														50	83	116	149	182	216	249	282	317	348	420	443	476	
Turkey	4	1	1	1	2	32	3	4	19	23	9	1	0	242	247	252	258	263	269	274	280	319	422	428	433	427	327
Uruguay	0		1	1	1	0	8	7	11	24	40	7	1	49	85	120	156	191	227	262	298	328	368	458	489	524	389
Greece	0	0	0	0	0	0	0	0	3	7	75	15	0	121	146	171	196	221	246	271	296	354	379	450	465	490	444
Japan	0	0	0	0	0	0	0	0	0	0	100	0	0													534	534
USA	0	0	0	0	0	0	0	0	2	31	61	6	0	60	99	138	177	217	256	295	335	379	458	498	509	548	484