



United Nations  
Educational, Scientific and  
Cultural Organization



UNESCO  
INSTITUTE  
FOR  
STATISTICS



GLOBAL  
ALLIANCE  
TO MONITOR  
LEARNING



# Profiling the global distribution of literacy skill

TCG Fifth Meeting  
15-16 November 2018  
Mexico City, Mexico  
**TCG5/ REF/9**

The paper was prepared by T. Scott Murray, DataAngel Policy Research, and Richard Desjardins, Stanford University, for the Global Alliance to Monitor Learning (GAML).

## High-level summary

- Estimates of the distribution of literacy skill by proficiency level are needed for a range of national and international purposes, including monitoring progress towards Sustainable Development Goal 4.6.1
- Given that a large number of countries are unlikely, or will be unable to field PIAAC and that proxy measures based on years of education are unreliable, alternative approaches are needed.
- Four options are described, three of which are based upon applying the relationships between skill and background characteristics observed in PIAAC to a limited matrix of national variables to yield synthetic estimates of the distribution of proficiency by level. These three options differ in the number of variables that are controlled for in the modeling and, thus, in the potential bias in the estimates. These options support the international need for data for monitoring purposes but only a limited subset of national needs. All three options reduce the financial, technical and operational demands of data acquisition significantly and yield estimates that are less biased than proxies based upon years of schooling.
- The fourth option involves fielding a PIAAC-based assessment that yields direct estimates of the probability of being at each proficiency level for a limited matrix of variables. This option supports virtually all of national and international data needs but imposes a lower financial, technical and operational burden than a full PIAAC implementation.
- None of the options improve the amount of measurement in the lower regions of the proficiency scale so will fail to detect much progress in countries with the largest proportions of their population in the lowest PIAAC proficiency levels.

The following table summarized the options presented and their pros and cons.

**Table 1: Options for generating literacy proficiency estimates for monitoring SDG 4.1.6**

	Options	Description	Pros	Cons
1	Field PIAAC/STEP	This option would see a large number of countries fielding the PIAAC assessment.	Yields fully comparable results for the full proficiency distribution, including lower levels	Expensive Technically demanding Operationally demanding
2	Generate model based estimates of proficiency distribution for limited matrix of characteristics	Generate national estimates using relationships between a limited set of variables and literacy skill observed in the PIAAC data set for a limited matrix of variable i.e. age group, gender, educational attainment and GDP per capita	Virtually costless allows countries to be grouped on probable level of need. Can be updated annually as new demographic estimates become available. Can be revised as new PIAAC data become available. Based on the testable assumption that the estimates support better policy decisions than years of education proxies	Will overestimate proficiency in lower regions of the distribution as a result of differences in initial educational quality, but less so than relying on differences in years of education
3	Generate model-based estimates of proficiency distribution for expanded matrix of characteristics	Generate national estimates using relationships between a limited set of variables and literacy skill observed in the PIAAC data set for an expanded matrix of variables shown to predict differences in skill. These variables would have to be added to an existing data collection.	Modest cost to collect additional variables such as employment status, occupation and skill use. Allows countries to be grouped on probable level of need. Can be updated annually as new demographic estimates become available. Can be revised as new PIAAC data become available. Based on the testable assumption that the estimates support better policy decisions than years of education proxies	Will overestimate proficiency in lower regions of the distribution as a result of differences in initial educational quality but less so than Option 2
4	Generate model-based estimates of proficiency distribution using full PIAAC variable set	Generate national estimates using relationships between a limited set of variables and literacy skill observed in the PIAAC data set for an expanded matrix of variables shown to predict differences in skill. The full PIAAC background questionnaire would have	Approach would reduce bias in estimates significantly. Relatively low operational and technical burden. Based on testable assumption that the estimates support better policy decisions than years of education proxies. Can be updated as new data become available.	Significant cost to collect full PIAAC background questionnaire. Significant reduction in bias.

	Options	Description	Pros	Cons
		to be added to an existing data collection.		
5	Generate proficiency estimates based on focussed skill assessment	Generate national score distributions using relationships between the full set of PIAAC background variables and an adaptive assessment of literacy that yields individually reliable scores and reliable estimates of the probability of being at each proficiency level for the full matrix of variables. Requires fielding a representative probability sample in which probabilities of selection have been adjusted to support reliable estimates of the probability of being at a level.	Less expensive than a full PIAAC implementation. Supports full range of both national and international data needs. Eliminates bias associated with applying other countries relationships	Technically, financially and operationally demanding (but less expensive than a full PIAAC implementation.)

## Executive Summary

Comparative data on the distribution of literacy and numeracy is required to monitor Sustainable Development Goal 4.6.1. Importantly, the rate at which literacy and numeracy distributions evolve is low, so low, in fact, that estimates are only needed every five years to track progress towards established targets. More directly, levels of year-to-year change are likely to be below the detection threshold of all but the most expensive skill assessment systems.

The required comparative estimates are only available for a small number of countries, a situation that forces multilateral organizations, bilateral donors and most national governments to rely on unreliable indirect proxies, such as years of schooling, to inform their decisions. A fundamental assumption underlying this work is that investment decisions could be improved were more reliable estimates available.

Ideally, the need for monitoring data would be met through widespread participation in the OECD's PIAAC assessment. PIAAC generates precise and reliable proficiency estimates that can be safely compared. It is highly unlikely, however, that a very large number of countries will field PIAAC over the coming decade, so alternative data will be needed for a large number of countries. Even where PIAAC data is available, quin-quennial estimates will be needed to be generated by some means for comparative purposes.

This paper sets out four options for filling these data gaps in an expeditious and affordable way.

The first option is to model the national distribution of literacy using the probability of being in each proficiency level observed in the OECD's PIAAC assessment for a fixed matrix of variables that are available for all countries i.e. age, education and gender and level of economic development. This approach would yield more reliable estimates than estimates based solely on years of schooling but would likely overestimate the proportions of adults in higher proficiency levels in less economically developed nations. These annual estimates could be generated rapidly at low cost using updated demographic estimates.

The second option would see the annual production of a set of global estimates of national literacy skill distributions generated by applying the relationships between proficiency level and a broader range of demographic variables included in the PIAAC data set. These variables would be selected to capture as much additional variance in skill as possible and the questions added to existing occasional data collections such as MICS, DHS or a recurring annual national sample survey. Estimates could be updated annually by applying the same model-based methods applied in option 1. This option would be significantly less expensive, operationally and technically demanding than fielding a full PIAAC assessment.

The third option would see production of baseline national literacy skill distributions generated by applying the relationships between proficiency level and the full PIAAC background questionnaire. The relationships of interest – the probability of being at a proficiency level for different groups of adults – would be obtained by administering an actual adaptive skill test that yields individually reliable estimates of skill, and the PIAAC background

questionnaire, to a small purposively selected sample designed to capture as much variance in skill as possible. The observed relationships would then be imputed onto a microdata file, such as the Census file, to generate estimates that are relatively free of sampling error. Estimates could be updated annually by applying the same model-based methods applied in options 1 and 2.

Given their likely distribution of skill, the model-based approaches are unlikely to detect much improvement in the proportions of adults in the lowest proficiency levels in the least developed countries.

Repeating this approach every 5 years would provide estimates based on evolving relationships. This option would be less expensive, operationally and technically demanding than the full PIAAC assessment.

## 1. Introduction

The UN Sustainable Development Goal 4.6.1 specifies a need to monitor the global supply of adult literacy and numeracy skill. The comparative data needed to support this requirement is currently available for a relatively small number of countries and is unlikely to grow at a rapid enough rate to support meaningful comparison. This report recommends options for filling the need to monitor the global supply of adult literacy and numeracy skills by other means.

More specifically, the report explores the adequacy of producing model-based estimates of literacy skill distributions for countries that have yet to field a reliable direct assessment of adult skill.

Such a set of estimates would not be able to satisfy the full range of data needed by public policy makers but would be more reliable than the indirect proxies that are currently being used as proxies for literacy skill such as years of education.

Such a set of estimates would, however, be reliable enough to serve as a baseline against which progress could be monitored at the international level.

Interested readers may contact the authors at the following coordinates:

T. Scott Murray  
DataAngel Policy Research  
1 Patterson Avenue  
Ottawa, Ontario  
Canada, K2L 2N9  
Email: [dataangel@mac.com](mailto:dataangel@mac.com)  
Mobile: +1 613 240 8433  
Web: [www.dataangel.ca](http://www.dataangel.ca)





## 2. Background

At a general level, data on the distribution of adult literacy and numeracy skill are needed to serve a broad range of national and international uses, including:

- For understanding the factors that determine the supply of literacy and numeracy skill and the impact that literacy skill have on individual, institutional and macro-outcomes.
- For establishing policy priorities and for designing efficient and effective policy and program responses
- For monitoring progress towards stated goals or targets and adjusting policy accordingly
- For evaluating the efficiency and effectiveness of policy and programs
- For administering literacy and numeracy programs including taking administrative decisions related to individuals, training providers and regions

The Sustainable Development Goals stipulate a specific need to monitor the adequacy of adult literacy and numeracy skills at the international level for the purposes of SDG 4.6.1.

These data are needed by multilateral and bilateral donors, and national governments, to allocate resources and to adjust these allocations in response to change in skill distributions over time.

In a perfect world, this need would be met through the implementation of a global skill assessment program that serves the full range of uses for which data is required. Such systems offer several advantages including:

- Amortizing fixed development and administrative costs over a large number of countries

Facilitating the implementation of quality assurance processes that reduce the risks of error and bias impairing the ability to make comparisons of skill distributions within and between countries over time.

- Providing explicit empirical evidence of comparability of results across key sub-populations

Analysis of the world's first cycles of international comparative adult skill assessment<sup>1</sup> confirm the basic premise underlying the studies i.e. that direct assessment of skill reveals far larger differences among countries, and among population subgroups within countries, than expected given differences in educational attainment. More directly, direct assessments

---

<sup>1</sup> Including data from the 1994 and 1996 International Adult Literacy Surveys (IALS), the 2003 and 2005 Adult Literacy and Life Skills Surveys (ALL) and the 2012 and 2016 Programme for the International Assessment of Adult Competencies (PIAAC)

provide much more reliable data for policy purposes than indirect measures of skill differences such as years of schooling or educational attainment, an important attribute given those differences in literacy skill explain over 55% of differences in rates of labour productivity and GDP growth over the long term (Coulombe, Tremblay and Marchand, 2007).

To date roughly 50 countries have fielded a direct assessment that is designed to support reliable comparison. Another smaller group of countries have fielded national literacy assessments that purport to support comparison within countries but do not support comparison across countries. For the most part, countries fielding direct assessments of skill have been drawn from among the world's most economically advanced nations.

Unfortunately, the balance of countries have yet to field a direct assessment of skill, a situation that can be traced to the fact that:

- National governments and donors have not made the needed budget allocations. By any standard such assessments are expensive,
- The available tools are not well suited to the needs of developing countries. Most importantly, it has been argued that the available studies do not provide sufficient discrimination in the lower ranges of the proficiency scales,
- The technical demands of implementing a direct assessment of adult skill exceeds the capacity of many developing countries, something that translates into high risks of implementation error.

This state of affairs that obliges multilateral donors, bilateral donors and a large number of countries national governments to rely on a mix of indirect skill measures, such as years of education or self-evaluation of skill, in setting their policy priorities and funding allocations related to adult literacy and numeracy.

On the surface, this seems a reasonable approach as differences of educational attainment explain roughly 60% of observed variance in literacy scores.

An extensive body of research suggests, however, that years of schooling and measures of educational attainment are relatively poor predictors of average skill levels, and of distributions of skill by proficiency level, by extension of differences among populations and population subgroups<sup>2</sup>.

Research traces this unreliability, in part, to significant differences in the quality of initial education. The balance in the variance in adult literacy scores is the product of skill gain and loss occurring over the life course. More directly, the variance in skill observed at the point of secondary graduation tends to amplify through processes that precipitate skill gain and loss. Adults who participate in adult education and training, who face high levels of non-routine skill use on the job, who high levels of non-routine skill use in private life and stable employment tend to increase their skill level. Adults who experience unemployment and low

---

<sup>2</sup> See for example Neice and Adsett, UNESCO Institute for Education's Seminar on Functional Adult Illiteracy, 1989

levels of skill use in their jobs and more generally in their private lives tend to lose skill with time. At a higher level, populations that generate higher levels of GDP per capita tend to have higher average skill levels and fewer adults in the lowest proficiency levels.

In general, adding in controls for age and gender and GDP per capita add an additional 15% - 20% of explained variance in skill, in large part because they capture some of the net skill gain and loss occurring over the life course.

Virtually all of the remaining variance in literacy scores can be explained by adding in controls for additional variables. Principle components analysis of data from the IALS, ALL and PIAAC adult skill assessment cycles reveals that the full set of questions carried on the background questionnaire fielded by these assessments capture up to 96% of observed score variance<sup>3</sup>.

With these facts in mind, the next chapter of this report sets out a pragmatic and ordered set of options that would address the need for more reliable data on literacy skill differences among countries to support national and international priority setting.

These options are designed to answer a subset of linked questions that are central to setting international and national policy priorities. The approach supports a broad range of reporting options, including the specification of the fixed proficiency thresholds required by SDG 4.6.1.

For international agents these questions include:

- *Which countries are most in need of assistance given their current skill distribution?*
- *Are skill levels rising at the expected rate? Which countries are in need of additional support to increase their rate of improvement?*

For national governments:

- *How do our skill level and distribution compare to our peers?*
- *Are skill levels improving at the expected rate?*
- *Are additional investments needed to meet social and economic goals?*

---

<sup>3</sup> See the International Adult Literacy Survey technical Report, US Department of Education, 1995



### 3. Proposed approach

The background set out above suggests that a pragmatic approach is called for, one that provides a menu of options from which multilateral organizations, bilateral donors and national governments might choose.

Three options are proposed:

1. Annual production of a set of global estimates of national literacy skill distributions generated by applying the relationships between proficiency level and education level, age, gender and GDP per capita observed in the PIAAC data set.
2. Annual production of a set of global estimates of national literacy skill distributions generated by applying the relationships between proficiency level and a broader range of variables included in the PIAAC data set. These variables would be selected to capture additional variance in skill and the questions added to existing data collections such as MICS and DHS.
3. Production of baseline national literacy skill distributions generated by applying the relationships between proficiency level and the full PIAAC background questionnaire. The relationships of interest – the probability of being at a proficiency level for different groups of adults – would be obtained by administering an actual skill test and background questionnaire to a small purposively selected sample designed to capture as much variance in skill as possible. The observed relationships would be imputed onto a Census file to generate estimates that are relatively free of sampling error. Estimates could be updated by applying the same model-based methods applied in options 1 and 2.

Each of these options is detailed below.

#### **Option 1: Model-based estimates**

This option involves the analysis of PIAAC national data sets to reveal the relationship between being at a literacy proficiency level and a small matrix of variables that both predict literacy skill and that are available annually for all countries in the world. These variables include:

- Education
- Age
- Gender
- GDP per capita

More specifically, each country in the PIAAC database would be analyzed to reveal the probability of being at each PIAAC proficiency level for each cell in the matrix defined by age, gender, education and GDP per capita.

The pooled estimated probabilities will be applied to aggregate estimates of the same data for countries lacking PIAAC countries and for countries for which PIAAC estimates are available. Comparison of the former estimates will be compared to estimate a table of deviations in the proportions at each proficiency level. This table will be analyzed to determine if deviations are sufficiently large to imply a change in the policy response.

Provided that the approach yields results that are deemed to be fit for some uses, the estimated relationships would be applied to aggregate estimates of the demographic matrix to obtain a likely national distribution of literacy skill by proficiency level for every country.

Of note, the approach would not impute scores, so it will not be possible to rank countries based on their average score. It will be possible to rank countries based on their proportions below a specified threshold such as the proportion at Level1 and 2 on the PIAAC scale.

By definition, these model-based estimates will be imprecise as none of the models capture the full variance in scores. Literacy data from PISA for 65 countries reveals that skill at the point of secondary graduation is far more variable than observed in the PIAAC data for 16 year olds. This observation suggests that the distribution of literacy by proficiency level will be biased upward for less developed countries.

Analysis of Canadian PIAAC data suggests, however, that skill gain and loss occurring over the life course is likely to attenuate much of the bias<sup>4</sup>. Some adults with low levels of educational attainment gain skill without the benefit of formal education and some adults with high levels of formal education lose skill as a result of low levels of use that itself reflects insufficient levels of economic and social demand to maintain skill stocks. Between 2003 and 2011 Canada shed roughly 7 points per adult as a result of skill loss during a period when average years of education rose a whole year, an increase that would have been expected to raise the population average skill level by 8 points. It is reasonable to assume that the same processes are influencing the supply of literacy skill available in developing countries even more.

No matter how imprecise and biased the model-based estimates are, it is important to remember that they will be more reliable than estimates based solely on distributions of educational attainment or year of schooling since they capture some of the otherwise unobserved variance in educational quality, net skill gain and loss and the causal relationship between skill levels and economic growth.

At the same time, these estimates are, by definition, error laden because they are based on inferred relationships rather than observed relationships.

The key underlying policy question is “Are the true relationships different enough to render the results unfit for use?” The answer to this question is most likely no except in cases where

---

<sup>4</sup> See **Reconstructing the Evolution of the American Supply of Cognitive Skills: A Synthetic Cohort Analysis**, a report by DataAngel Policy Research for the US National Center of Education Statistics, 2015, **Understanding Literacy Markets in Alberta: A Segmentation Analysis**, DataAngel Policy Research, 2015 and **Literacy in Canada: Trends in skill supply and skill demand, skill shortages and surpluses and skill loss**, DataAngel Policy Research for the Council of Ministers of Education Canada, 2018

extended conflict has disrupted the education system and the economy. More precisely, individual percentile ranks might change but countries are unlikely to shift decile ranks.

A second comparison estimated proportions at proficiency levels would be made that involves comparing the distributions of countries that participated in the World Bank STEP and UNESCO's LAMP assessments.

In the case of STEP the estimates will differ by design as the STEP assessments generally only reflect the urban population in assessed countries. In this sense the model-based estimates can be used to judge the magnitude of bias associated with restricting STEP coverage.

In the case of LAMP, the estimates of the distribution of proficiency by level could be compared directly if an linking study was undertaken that would use the ALL prose literacy and document literacy items included in the LAMP item pool. This comparison would provide a sense of the degree to which international pooled predictors mirror the relationships observed in the LAMP dataset.

### **Option 2: Modeled estimates of literacy skill distributions by proficiency level based upon a broader set of predictors**

The predictive power of the models applied in Option 1 could be improved by including additional variables in the imputation matrix that have been shown to have predictive value.

These variables, inter-alia, include language of instruction/mother tongue, employment status, occupation group, indices of skill use on the job and in private life, immigrant status and participation in adult education and training.<sup>5</sup>

Some of these variables may already be available from other sources such as the DHS or MICS. In this case an expanded predictive demographic matrix would be estimated using the PIAAC data and an expanded predictive matrix tabulated from the secondary source. It is expected that the additional variables will add between 15% - 20% to the explained variance so are expected to improve the reliability of the predicted distributions.

Variables that are not available could be added to on-going data collections that are based upon fielding representative samples of adults aged 16 to 65 and the required demographic estimates tabulated.

### **Option 3: Conduct of a focused literacy assessment**

As recommended in previous reports produced for UIS, a third option that might be implemented is for additional countries to field a focused assessment of literacy skill that is specifically designed to estimate the probabilities of being at each proficiency level rather than one that attempts to estimate scale scores.

Such a study would still use a representative sample of adults but the probabilities of selection would be adjusted to maximize the amount of variance captured in the model of predicting

---

<sup>5</sup> See Desjardin's PhD thesis, University of Stockholm, 2003

the distribution of proficiency levels. In general, samples of roughly 1,000 – 1,200 would be required, a number that yields a significant reduction in operational and financial burden when compared to fielding a full assessment such as PIAAC.

Such a study would also benefit from fielding an assessment that yields scale scores that are individually reliable. The population-level assessments that have been conducted to date have all been designed to yield reliable estimates for population subgroups, a feature that reduces the number of test items that need to be administered. A fully adaptive test would yield individually reliable proficiency estimates and would, thereby, improve the predictive value of background covariates significantly. Fully adaptive tests reduce the number of test items needed by roughly 40% so yield concomitant reductions in cost and burden of roughly the same magnitude.

The data from the focused assessment would be analyzed with regression techniques to reveal the relative impact of background variables on scores.

The estimated regression parameters are then used to impute literacy scores on the 500-point proficiency scale that are then grouped into proficiency levels.

Although scores could be imputed onto any representative data set, ideally scale scores would be imputed onto a Census micro-file. This approach would greatly reduce sampling error associated with the estimates, leaving standard errors dominated by imputation error. Provided that the variables included in the model capture most of the variance in scores, national estimates will be reasonable proxies for true distributions.

An additional advantage of this approach is that it produces a dataset with the full set of PIAAC background questions and reliable scores that can be used to profile the distribution of skill, analyze the determinants of skill differences among population subgroups and to analyze the impact of skill on a broad range of education, labour market, health and social outcomes, albeit at a less detailed level than supported by a full PIAAC implementation.



## 4. Summary, conclusion and recommendations

Data on the distribution of literacy skill, the variables that predict differences in literacy scores and proportions at proficiency levels and the impact that literacy skill has on individual outcomes is needed by a broad range of users and uses. This need is formally embodied in the Goal 4.6 of the Sustainable Development Goals that specify a need for all adults to have adequate levels of literacy and numeracy skill.

Ideally, the needs of all users and uses, including the SDG's, could be met through participation in an international comparative assessment of adult skill such as the OECD's PIAAC.

Unfortunately, a relatively small proportion of the world's countries have fielded such an assessment in part because participation imposes a significant technical, operational and financial burden.

Three facts suggest literacy skill is fundamental to economic and social development.

First, differences in literacy skill have been shown to explain 55% of differences in long term rates of labour productivity and GDP growth, something that makes them the most important determinant.

Second, differences in literacy skill underlie most inequality in valued educational, labour market, health and social outcomes

Third, weak literacy skills reduce the productivity of public investments in education and health.

The absence of comparable estimates of skill obliges several key users – multilateral organizations, such as UNESCO, UNICEF, WHO, ILO, the World Bank, the regional development banks, bilateral donors such as USAID, DFID, GTZ, SIDA and NORAD, and national governments with data to answer key questions including:

- *What is the current distribution of literacy skill by proficiency level?*
- *How do literacy skill distributions compare? Who has the greatest need for support?*
- *Is literacy proficiency rising rapidly enough to meet social and economic objectives? What additional investment and policy and program adjustments are needed to do so?*

Make no mistake – these organizations will still make decisions on investment priorities and funding allocations in the absence of reliable data on relative importance and need. By definition, these decisions will lead to sub-optimal investment decisions, something that developing countries can ill afford.

This report recommends three options that would provide this subset of policy makers with comparative information on literacy skill that is needed to make better decisions about investment priorities and resource allocation.

The first option involves generating a global set of estimates of the distribution of literacy skill by proficiency by applying the relationships among age, gender, education and GDP per capita observed in pooled PIAAC data to aggregate national data for these same variables.

The validity of this approach is tested by comparing the estimated proficiency distributions to the distributions observed in each PIAAC dataset. It is expected that these variables will capture roughly 70% of variation in skill differences.

The second option involves expanding the number of background variables included in the regression/imputation model. Variables would be selected from the PIAAC background questionnaire that have been shown to predict differences in literacy skill distributions. These variables include, inter-alia, mother tongue, employment status, occupation group, participation in adult education and training and indices of skill use on the job and in private life. In some cases, some of these variables may be available from existing sources. Where variables are not available from existing data sets these variables would be added to the next cycle of data collection.

It is expected that these variables will capture roughly 80% of variation in skill differences.

The third option would be to run a focused assessment and background questionnaire that is designed to yield estimates of the probability of adults in a limited demographic matrix of being at each proficiency level. The requisite probabilities would be estimated from a representative sample of roughly 1,000 to 1,200 cases in which the probabilities of selection have been adjusted to capture as much variance in score predictors as possible. The sampled background characteristics are likely to capture 95% of the variance in scores. The estimated probabilities would be used to impute scores onto the most recent Census micro file. This option would preclude the need for model-based estimates generated by options 1 and 2 and, within the limits imposed by the sample size, will support most of the policy and program design uses to which assessment data is put.

## Annex. Predicting national literacy skill distributions

### Introduction

Sustainable Development Goal 4.6.1 stipulates a need to monitor progress towards a goal of the adult population having functional literacy and numeracy skills.

A small minority of countries have access to the direct measures of adult literacy skill that are needed to support comparison of skill distributions within countries or to monitor progress over time.

The balance of countries rely on unreliable indirect measures of literacy skill distributions.

This report documents the results of a statistical analysis designed to yield more reliable estimates of literacy skill distributions for these latter subset of countries.

Interested readers may contact the authors at:

T. Scott Murray  
DataAngel Policy Research Incorporated  
1 Patterson Avenue  
Ottawa, Ontario K2L 2N9  
Canada

Email: [dataangel@mac.com](mailto:dataangel@mac.com)  
Web: [www.dataangel.ca](http://www.dataangel.ca)  
Mobile: +1 613 240 8433

### 1. Background and methods

Currently, the overwhelming majority of the world's countries rely on indirect measures of literacy skill that have been shown to be both unreliable and incomparable. These indirect measures include estimates inferred from distributions of years of schooling, estimates inferred from self-assessments of adult literacy skill and from the administration of very simple reading tasks.

Direct measures of adult literacy skill for roughly a quarter of the world's countries. These direct measures are obtained by testing representative samples of adults with a literacy test that is designed to cover the full range of proficiency. Direct measures are designed to support reliable comparison within and between countries and over time. The estimates produced from these direct assessments – IALS, ALL, PIAAC, STEP and LAMP - are reliable enough to support most intended uses and, despite being collected at different points in time, support comparison within and between populations, and over time.

The direct assessment of literacy skill is financially, technically and operationally demanding. As a result, the number of countries with access to direct measures of literacy skill It is unlikely to growth rapidly enough to meet the demands of SDG 4.6.1.

This paper documents the results of an analysis that is designed to approximate literacy skill distributions by using the relationships between skill and a limited set of demographic characteristics observed in the PIAAC data set. More specifically, the estimated probabilities of being at each proficiency level observed in the PIAAC dataset for a limited set of demographic characteristics are applied to a current set of estimates of the same demographic characteristics at the national level.

The available research on the determinants of differences in literacy skill among populations suggests several things, including that:

1. Estimates of literacy based upon years of schooling are unreliable because they fail to reflect any of the variation in educational quality among population subgroups with countries or between countries, nor the differential skill gain and loss among population subgroups with countries or between countries. Analysis of the IALS, ALL, PIAAC and STEP datasets suggests that years of schooling only capture roughly 60% of the observed variation in scores among countries.
2. Estimates of literacy based upon self-reports are unreliable because they are biased by the individual's context. More specifically, they are conditioned by the skill levels of the people with whom the individual interacts on a regular basis and upon the level of economic and social demand that they face on a daily basis. In general, individuals in tend to overestimate their true skill level.
3. Estimates based upon the administration of a single reading task are unreliable because they do not systematically sample the variables that underlie differences in task difficulty, and, by extension, proficiency. They are also biased by differences in the cultural familiarity with the content among groups.
4. Estimates based upon the available international comparative direct assessments – IALS, ALL, LAMP, PIAAC and STEP – provide empirical confirmation of their reliability and comparability. This said, the PIAAC assessment offers very little measurement of the lower regions of the literacy proficiency scale, something that limits one's ability to identify groups sharing common learning needs.
5. Estimates based upon the available international comparative direct assessments – IALS, ALL, LAMP, PIAAC and STEP – have also been shown to be predictable, in the sense that differences among population subgroups with countries and among countries have been shown to reflect differences in the background characteristics of respondents. The full set of PIAAC background characteristics capture roughly 96% of the observed variance in literacy skill in the IALS, ALL, PIAAC and STEP datasets.

The proposed approach estimates the probability of individuals sharing a common set of demographic characteristics defined by:

Education (8 categories);

Age (4 categories),

Immigration-language status (4 categories);

Parents' education (3 categories);  
 Labour force status (2 categories);  
 Occupation (4 categories);  
 Earnings (5 categories);  
 Reading at work index;  
 Reading at home index.

The first 7 of these variables are readily available for virtually every country in the world on a current basis. Collectively, these variables capture roughly 80% of variance observed in the IALS, ALL, PIAAC, LAMP and STEP databases.

The last pair of variables – the reading and work and reading at home indices – capture an additional 10% of observed variance. These data are not currently available for most countries but could easily be added to a household survey to provide the required estimates.

The model also adjusts the estimated probabilities for different types of countries based upon where they fall in the distribution of average literacy scores, their GDP per capita and their level of income inequality. The groups, referred to as tiers in the analysis, are defined as follows:

TIER 1- Lower scores and education, lower GDP, higher inequality (gini>40) (larger shift from Level 2,3,4,5 to Level 1 or below) (e.g. Turkey, Chile)

TIER 2- Lower scores and education, medium-lower GDP, higher inequality (gini>40) (smaller shift from Level 2,3,4,5 to Level 1 or lower) (e.g. Israel, Singapore)

TIER 3- Lower scores and education, medium-lower to medium-higher GDP (28000-38000 GDP), medium-high to higher inequality (Gini 35-40) (smaller shift from Level 3,4,5 to Level 2 or lower) (e.g. Spain, Greece, Italy)

TIER 4- Medium scores and education, medium-lower to higher GDP (32000-50000 GDP), lower to medium inequality (gini 26-35), (shift from Level 4/5 and Level 1 or lower to Levels 2 and 3) (e.g. Germany, Denmark, Ireland...)

TIER 5- Medium to higher scores and education, medium-higher to higher GDP (43000-54000 GDP), lower to medium inequality (gini 26-36) (shift from Level 2 or lower to Level 3,4,5) (e.g. Finland, Netherlands...)

Higher proportions of adults in PIAAC Below Level 1, Level 1 and Level 2 literacy have been shown to reduce rates of GDP and labour productivity growth and to increase the potential for high levels of skill-based employment, wage and income inequality (Coulomb, Tremblay and Marchand, 2004; Coulombe and Tremblay, 2007 and Wiederhold, S. and Schwerdt, G., 2018). Thus, the addition of these variables captures some of the otherwise unobserved variation in scores attributable to differences in the effectiveness of education systems and in how skills get utilized in the economy to generate output. It is difficult to estimate how much additional variation in scores the addition of these variables captures.

Given the evidence presented above, it is reasonable to deduce that the proposed model-based estimates will outperform estimates based upon years of schooling, self-reports or the administration of a simple reading task.

The resulting national literacy distributions are more reliable than indirect measures of literacy but yield slightly different distributions of literacy skill by proficiency level for high-income countries. Indirect measures are likely to over-estimate skill levels.

To be fit for use whatever estimates are generated must be close enough to the true distributions to support several key inferences including:

*Which countries have the greatest need to improve their literacy skill level?*

*Which countries are failing to improve their literacy skill distribution at the expected rate?*

## 2. Results

In order to evaluate the reliability of the model-based estimates the estimated parameters were used to predict the distribution for the PIAAC countries for which PIAAC data is available. Tiers were created in order to approximate a type of country so that the relationships between literacy skill and background variables applied most closely reflect the expected relationship in the target country.

The full results for 31 PIAAC countries are provided in the attached Excel file:

UIS literacy model -master.xlsx

UIS data on educational attainment by gender was used to predict the distribution of literacy skill by proficiency level in the United States of America. The predicted distribution for the USA was:

Level 1 or below=5.3 (actual 4.6%);  
Level 1=11.9% (actual 14%);  
Level 2=26.5% (actual 33.6%);  
Level 3=36.1% (actual 35.1); and,  
Level 4/5=20.2 (actual 12.6%).

In short, in the case of the United States, the model using only educational attainment by gender:

- slightly over predicts Level 1 or below by only 0.7 percentage points
- under predicts Level 1 by 2.1 percentage points
- under predicts Level 2 by 7.1 percentage points
- over predicts Level 3 by 1.0 percentage points
- over predicts Level 4/5 by 7.6 percentage points.

Expressing the error as the difference divided into the actual (a sort of coefficient of variation) the results in order are:

+13.9%,  
-18.1%,  
-27.0%;  
+2.9%  
+37.6%.

It is clear from the foregoing results that the most parsimonious model – the one that uses only educational attainment by gender – are quite unreliable.

The US has high levels of educational attainment and so more people should score at Levels 3, 4/5. But because of skill depreciation, other factors, and schooling quality issues (none of which can be predicted from UIS education data), more people actually score at Level 2 and 1 in the US than would otherwise be predicted on the basis of attainment data alone.

One would expect the opposite to be true for countries with very low levels of educational attainment i.e. there will be a tendency to over predict low levels of literacy for those with very low attainment levels. For example, Italy does a lot better in PIAAC than would otherwise be predicted from attainment data —probably a function of its advanced economy.

TIER adjustments for the other countries, such as Finland, Italy, Chile and Turkey, that introduce national adjustments by proficiency level, improve the reliability of the estimates slightly. Estimated CV errors range from: 9.6-41.4% for Finland, 1.1-66.2% for Italy; 2.7-40.9% for Chile – large enough to alter the policy prescription that one would infer from the results.

Predictions for non-OECD countries - Bolivia, Mexico, Saudi Arabia and Vietnam –yield plausible results that are no doubt subject to significant upward bias because they assume that these countries have the same average educational quality as more advanced countries of the OECD.

Data from PISA, TIMSS and PIRLS suggests that this assumption does not hold i.e. that educational quality is much lower in lower income countries. In addition, nothing is known about the impact of net skill gain and loss in less developed countries on the levels and distribution of skill. In those countries of the OECD where PIAAC data is available for multiple assessment cycles, skill loss occurring over the life course is large enough to more than offset skill gain realized by rising education levels.

The only way to improve the reliability of the predictions is to include more variables in the model that have been shown to capture additional variance in observed skill distributions.

Work by one of the authors (Murray) outlines options for collecting the required variables needed to improve the models. The same work proposes that a modified PIAAC design that would provide a full PIAAC skill/background covariance matrix while reducing the financial, technical and operational burden of fielding the full PIAAC assessment.

### 3. Summary and conclusion

Policy makers currently rely on a range of unreliable methods to estimate the distribution of adult literacy skills by proficiency level for the majority of countries in the world.

The methods employed for these countries are not able to meet the need for reliable and comparable estimates of adult literacy needed to monitor progress of all countries towards the targets specified in Sustainable Development Goal 4.6.1.

Methods that yield reliable and comparable data are available but are currently only available for a relatively small subset of countries that have fielded a comparative assessment of adult literacy skill.



There is little prospect that the number of countries having fielded a direct assessment of literacy skill will grow rapidly enough to fill this data gap. Thus, there is a need to meet the data need through alternate means.

This paper explored whether the data need could be met through alternate means. Specifically, the paper evaluates whether applying the relationship between literacy skill and various subsets of the demographic variables collected on the PIAAC background questionnaire to an existing dataset could yield estimates that are fit for the SDG use.

The analysis undertaken suggests that the options explored empirically do not yield estimates of the required precision. More specifically, the model that includes only educational attainment and gender yields a distribution of literacy by proficiency level that is sufficiently different from the actual assessment-based distribution to render it unfit for use.

The inclusion of additional background variables would improve the predictive validity of the model but would require countries to obtain additional background variables, minimally including:

- Age (4 categories),
- Immigration-language status (4 categories);
- Parents' education (3 categories);
- Labour force status (2 categories);
- Occupation (4 categories);
- Earnings (5 categories);
- Reading at work index;
- Reading at home index.

Where these variables are not already available from a secondary source the variables would have to be appended to an existing household survey.

The inclusion of these variables in the model would increase the amount of controlled variance to roughly 80% - 90% and would, thus, provide estimates that are fit for use in monitoring SDG 4.6.1.

Collecting the entire PIAAC background questionnaire would improve the quality of the model-based estimates as these variables capture roughly 96% of observed variance in literacy scores.

An alternative would be to field a direct assessment of literacy in which the sample design and allocation has been adjusted to yield reliable estimates of the probability of being at a proficiency level that could be used to generate model-based estimates. This approach would reduce the financial, operational and technical burden of fielding a full skill assessments such as PIAAC, STEP or LAMP and yield estimates that reflect idiosyncratic relationships in every country.

