



United Nations
Educational, Scientific and
Cultural Organization



UNESCO
INSTITUTE
FOR
STATISTICS



GLOBAL
ALLIANCE
TO MONITOR
LEARNING

GAML4/REF/15

Functional literacy and numeracy: Definitions and options for measurement for the SDG Target 4.6

November 2017

DRAFT NOT FOR DISTRIBUTION

Abstract

The Sustainable Development Goals specify a need to measure the literacy and numeracy levels of the adult population 15 and over. This report recommends definitions for literacy and numeracy and proposes a strategy for monitoring progress.

Prepared by:
T. Scott Murray
DataAngel Policy Research Incorporated
dataangel@mac.com
www.dataangel.ca

Executive summary

The Sustainable Development Goal target 4.6: states that “By 2030, ensure that all youth and a substantial proportion of adults, both men and women, achieve literacy and numeracy.” The global indicator for the Target 4.6, and the only indicator for this target directly related to the measurement of learning outcomes, is the indicator 4.6.1, which states: the percentage of the population in a given age group achieving at least a fixed level of proficiency in functional (a) literacy and (b) numeracy skills. The target age group for this indicator is the population of 15 years and older.

This report recommends that the literacy and numeracy indicators be based upon the definitions of literacy and numeracy used in the OECD’s PIAAC adult skill assessment program. These definitions are precise enough to be measured and broad enough to capture the entire range of skill encountered globally. Although the PIAAC assessment was only administered to 16 to 65 year olds, the ALL assessment included all adults 16 and over. Analysis of results for the ALL countries confirms that the test works equally well in older populations.

The report also proposes a strategy for monitoring progress, one that offers countries a range of options.

Countries on their way to achieving universal secondary education are encouraged to participate in the next round of PIAAC data collection scheduled for 2021. The PIAAC design and processes are based upon 35 years of development and yield results that are valid, reliable, comparable and interpretable.

For countries below this level of educational development, the current PIAAC design offers a quite limited information return on their investment. Moreover, the technical, operational and financial burdens imposed by PIAAC may be too great for some countries to bear, something that translates into a non-negligible risk of catastrophic failure. Where this is the case, it is recommended that countries reduce the technical, operational and financial burden of PIAAC participation by:

Switching collection entirely to computer-based methods

Administering a fully adaptive test that provides equally reliable proficiency estimates over most of the two established PIAAC proficiency scales: literacy and numeracy

Employing purposive sampling to generate estimates of the probability of key population sub-groups being at each proficiency level and then using these probabilities to impute proficiency scores onto their 2021 Census of Population (DataAngel, 2012; CLLN, 2014). These synthetic proficiency estimates have been shown to be sufficiently precise and reliable to inform the full range of national and international policy questions through comparison to the PIAAC estimates.

Where countries are unwilling or unable to collect data in the current cycle, UNESCO should generate model-based estimates using the probabilities observed in comparator countries. This will ensure that measured progress reflects more than changes in the coverage of the countries for which data is available.



1. INTRODUCTION	4
2. THE CONCEPTUAL EVOLUTION OF LITERACY AND NUMERACY	5
3. THE USES AND USERS OF COMPARATIVE LITERACY AND NUMERACY ASSESSMENT RESULTS	7
4. DEFINITIONS OF FUNCTIONAL LITERACY AND NUMERACY	15
5. THE WAY FORWARD: RECOMMENDATIONS FOR DEFINITIONS AND MEASUREMENT	17
6. AVAILABILITY OF THE RANGE OF ITEMS FROM DIVERSE ADULT LITERACY ASSESSMENTS THAT COULD BE USED TO REPRESENT THE MEASURE OF ADULT FUNCTIONAL LIFE SKILLS AND COMPETENCIES AND COULD BE USED TO BENCHMARK MINIMUM PROFICIENCY LEVEL	25
7. IDENTIFICATION OF MEASUREMENT OPTIONS	28
8. RECOMMENDATIONS FOR THE WAY FORWARD: A SUMMARY	33
ANNEX A: REFERENCES	38
ANNEX B RECENTLY APPLIED DEFINITIONS OF LITERACY	39

1. Introduction

This report provides recommendations regarding the definition of literacy and numeracy that should be adopted in support of Sustainable Development Goal target 4.6: “By 2030, ensure that all youth and a substantial proportion of adults, both men and women, achieve literacy and numeracy.”

This objective is met through a review of the related literature, including a review of the definitions that have been employed in recent national and international attempts at measurement.

The report also proposes recommendations with respect to a measurement strategy that supports the goal. The global indicator for the Target 4.6, and the only indicator for this target directly related to the measurement of learning outcomes, is the indicator 4.6.1, which states:

“The percentage of the population in a given age group achieving at least a fixed level of proficiency in functional (a) literacy and (b) numeracy skills. The target age group for this indicator is the population of 15 years and older.”

This objective is met through a review of how the definition of literacy has evolved over the past three decades, of the criteria needing to be met to support the intended uses of the assessment data and how current approaches might be relaxed to reduce the technical, operational and financial burden without compromising the intended uses of the data.

2. The conceptual evolution of literacy and numeracy

Chapter 6 of the Education for All Global Monitoring Report of 2006 provides a useful summary of how the concept of literacy and numeracy has evolved significantly over the past 40 years.

In the 1970's UNESCO defined literacy indirectly by associating skill with years of formal education. More specifically, individuals were judged to be functionally literate if they attended 4 years of education and to be fully literate if they attended 9 years of education.

UNESCO also relied on self-reports of literacy and upon the administration of a simple reading passage in the context of the Census of Population.

Theoretical developments related to the definition and measurement of literacy and numeracy in the United States were applied in the 1985 US YALS, the 1987 Canadian LSUDA and the 1990 US NALS assessments. These developments advanced the notion of literacy and numeracy skill as continua rather than dichotomies, and documented that the relative difficulty of reading and numeracy tasks were largely a function of the cognitive demands of the task rather than the text or operation being undertaken. Being able to predict the relative difficulty of tasks a priori opened a way to systematically sample the determinants of task difficulty in the assessment design. Analysis of the data from these three studies revealed that all three of UNESCO's approaches to measurement were seriously flawed. More specifically, large proportions of adults who were classified as functionally literate or fully literate were, in fact, much less skilled. Similarly, large proportions of people with limited formal education had found a way to become highly literate.

About the same time, John Bynner and his colleagues undertook a longitudinal analysis of the literacy and numeracy skills of young Britons using the repeated assessments included in the British Birth Cohorts study. Bynner was the first researcher to document significant and rapid skill loss after the point of exit from the formal education system. This insight led, eventually, to the elaboration of practice-engagement theory (Reder, 2014) in which observed skill levels in adulthood are a function of skill gain and loss that is conditioned by the frequency and range of skill use undertaken. Literacy was no longer a static commodity obtained through education but rather a dynamic commodity – more like a muscle that required exercise.

Also about the same time Paulo Freire wrote *Paedagogy of the Oppressed*, a book that highlighted the central importance of literacy to individual empowerment and agency.

The elaboration of the "New Literacy Studies" conception of literacy suggested that the ways in which literacy is practised varies by social and cultural context. Ironically, a key insight offered by analysis of the international comparative data from YALS, LSUDA, NALS, IALS, ALL, ISRS, LAMP, PIAAC and STEP is that skill use varies a great deal both within and between populations with identical characteristics. What advocates of the "New Literacies" approach fail to apprehend is that the systems that condition adults practice are themselves amenable to change. More directly the PIAAC "markets" framework includes elements of skill supply and how it is generated, skill demand and how it evolves and skill utilization, how the market allows people to apply their skill fully and with what impact on their outcomes. Government policy choices can influence all three elements. The "New literacies" advocates also tended to confound how one acquires literacy and numeracy with the highly variable ways in which it is used in daily life. The latter dimension of use is both conditioned by the individuals objectively measured skill level and by the environment in which they find themselves. Whatever the recursive nature of this relationship having high levels of objectively assessed literacy and numeracy skill translate into much more positive individual outcomes that aggregate into positive macro-economic

and social outcomes. The key goal for policy makers has to be to ensure that all adults have the literacy and numeracy skills needed to be efficient and effective learners and that the social and economic systems that mediate skill use utilize the available skill supply fully.

Central to Fiere's pedagogy is the notion of 'critical literacy', a goal to be attained in part through engaging with books and other written texts, but, more profoundly, through 'reading' (i.e. interpreting, reflecting on, interrogating, theorizing, investigating, exploring, probing and questioning) and 'writing' (acting on and dialogically transforming) the social world. This notion is central to the notion of literacy and numeracy embodied in the PIAAC framework - as tools that facilitate the acquisition of information and its thoughtful application.

By 1978, UNESCO's General Conference adopted a definition of functional literacy - still in use today - which states: 'A person is functionally literate who can engage in all those activities in which literacy is required for effective functioning of his group and community and also for enabling him to continue to use reading, writing and calculation for his own and the community's development.'

Despite the apparent clarity of this definition a large number of competing definitions - reproduced in Annex B of this report - have emerged, each, purportedly, designed to support comparative assessment at the national or international level, This chapter has provided an overview of how the conception and definition of literacy and numeracy have evolved over the past several decades. Current theory provides a useful frame for thinking about how literacy is acquired, how it is used and how its use impacts individual and collective outcomes in the areas of the labour market, health, education, social and democratic domains. The next chapter narrows the view by offering a summary of uses to which comparative skill assessment data are put and who uses the results.

3. The uses and users of comparative literacy and numeracy assessment results

This chapter sets out the uses and uses of literacy and numeracy assessment results and reflects on what the intended uses and users imply for the design and reporting of said results.

At the highest level, systems that collect and disseminate official statistics serve a range of uses, including:

- Knowledge generation, a process of understanding causal relationships and effect sizes
- Program policy and planning, the development of public responses to identified problems
- Monitoring, a process that tracks progress against established goals and identifies the need for adjustment
- Evaluation, a formal process that analyses the efficiency and effectiveness of specific policy and program measures
- Administration, the process of making decisions about specific individuals or organizational units

To be useful the data collected in support of the Sustainable Development Goal 4.6 must serve the first three of these purposes.

It is possible that the tools used to assess literacy for the first three purposes can also be used to evaluate and administer programs, provided that variants of the assessment tools are available that are sufficiently precise to support these applications. In practice, such tools need to be significantly more precise so that misclassifications – either false positives or false negatives – are kept within acceptable minimums.

With this as context, it is important at the outset to reiterate the specific objectives that underlie the need for a global set of skill estimates that can be compared. Comparative assessments of literacy and numeracy yield three sets of key estimates:

- The average level of literacy and numeracy skill, and the distribution of literacy and numeracy skill by proficiency level, on proficiency scales that may be compared across countries. UNESCO views literacy and numeracy first and foremost as human rights – so knowing who has acquired what level of skill – is an important goal of comparative assessment. These data are also needed to understand competitive position and whether the skill supply is able to support national economic and social goals.
- The impact that literacy and numeracy skill have on individual labour market, health, educational, democratic and social outcomes. These data are needed to ascertain how much of social inequality is skill-based and this amendable to policies that raise skill levels.
- The variables that underlie or explain observed differences in average literacy skill levels and the distribution of literacy and numeracy. These data are needed to identify where policy might best be focussed and how rapidly skill levels might be improved.

The need for estimates of the level and distribution of skill is met through having results reported on a valid, reliably and comparably measured and interpretable international scale.

Validity rests upon a clear definition of the concept to be measured, theory that allows researchers to predict the relative difficulty to reading tasks over the range of difficulty expected in the target populations and is uni-dimensional. In reality analysis of the ISRS, LAMP, PIAAC and STEP reading components data shows that literacy is, in fact, multi-dimensional. In the lower regions of the proficiency scales adults are still in the process of mastering the mechanics of reading i.e. they are still in the process of learning to read. Once they have become fluid and automatic readers they can devote their cognitive resources to applying what they have read through analysis, integration, evaluation and synthesis i.e. they are able to read to learn. Most adults have achieved mastery of the requisite reading components by roughly 250 on the PIAAC literacy scales. Numeracy also displays some multi-dimensional properties in that adults must first read well enough to understand the nature of the question being asked and the nature of the expected response. Thus, numeracy skill is, by definition, highly dependent upon literacy skill.

Reliability rests upon the ability of the approach to measurement to detect meaningful differences in reading proficiency, that do not depend on the selection of items taken and on the availability of empirical evidence that the proficiency estimates can be interpreted as proxies for general proficiency.

Comparability rests upon the application of statistical methods that provide empirical evidence that the test supports comparison across population subgroups within and between countries, over independent samples and over time.

Interpretability rests with competence being measured in a valid and reliable way, that test scores are associated in a causal way with material individual outcomes and that skills, as measured by increases in test scores, could be increased through instruction.

It is worth noting that, for countries at roughly the same level of economic and educational development, the amount of variance in test scores within each population is far higher than the amount of variance observed between countries. This implies that any test capable of capturing the variance in test scores within heterogeneous national populations will be able to support international comparisons provided the empirical information flowing from the scaling confirms that the test is operating in a psychometrically stable way. The psychometric evidence accumulated through the conduct of NALS, IALS, ALL, ISRS, PIAAC, STEP, and LAMP confirms that their shared approach to measurement yields comparable results over the full range of proficiency observed in a diverse set of countries and languages. In the upper regions of the PIAAC proficiency scales performance depends on the application of universal cognitive strategies that engage the reasoning processes in the pre-frontal cortex. In the lower regions of the scales proficiency is related to mastery of a set of reading skills that themselves vary from language to language. For example, letter recognition in English is simpler than in Arabic because there are fewer symbols to memorize. Similarly, decoding in Spanish is less difficult than is English because every phoneme is voiced. In Chinese letter recognition and decoding are folded into receptive vocabulary skill.

The net result is that scores above 250 on the PIAAC proficiency scales may safely be compared directly and scores between 0 and 250 may be compared indirectly e.g. what proportion of a given population have mastered the reading components of their language whatever these might be. It is fundamentally important to note, however, that the addition of the reading components measures transforms the nature of the underlying proficiency scale from a quasi-interval scale to a true interval scale because one can define an absolute zero on the scale i.e. someone who cannot identify a single symbol. This transformation allows one to place people on the same, invariant scale despite the fact that the process by which people acquire literacy differs across languages.

The latter two needs – the need for data on the determinants of literacy and numeracy and their link to outcomes – are met through the administration of an internationally comparable background questionnaire. The current PIAAC background questionnaire explains a very high proportion of observed variation in test scores over the sampled populations.

More prosaically, the literacy and numeracy proficiency data is needed to provide national policy makers with answers to the following ordered set of policy questions:

- Do our skill levels support our social, economic and democratic goals?
- How do our skill levels compare to those of our key trading partners and neighbor countries?
- Are any identified skill gaps likely to impair our ability to compete and/or to meet our social and economic objectives?
- How well is the education system performing at generating new skill supply when compared to other countries? Is the education system able to supply enough skill supply to meet projected skill demands? Or to reduce the current level of skill shortage?
- Which population subgroups are most at risk of having inadequate skills?
- What kind of measures might be needed to reduce the size of any identified skill gaps? What would these measures cost? What is the likely return on these investments? Where would investments yield the highest rates of return in terms of reductions in skill-based social inequalities or improvements in economic outcomes?
- What is the level of skill-based inequality in labour market, health, educational, democratic and social outcomes? Are the observed inequalities large enough to pose a threat to social stability?
- What are the consequences of failing to reduce observed levels of skill shortage?

International data users need comparable estimates of literacy and numeracy to identify which countries have the greatest needs and where their loans, grants and bilateral aid would have the largest returns on investment.

It is worth noting in passing that the policy process into which the assessment results will be inserted do not require very precise estimates to be useful. For example, policy makers need only know that their rank, as judged by average scores, is roughly as expected to the average scores of key trading partners. More directly, a shift in rank of a few places is unlikely to alter the policy prescription much, if at all.

Similarly, comparisons of the distribution of literacy and numeracy skill by proficiency level will reveal that significant proportions of adults lack the skills to satisfy expected occupational demand, or to achieve national social and economic goals. The absolute magnitude of the “literacy” problem is unlikely to impact the policy prescription, just the magnitude of resources and time frame for remedying any identified skill shortages. More directly, whether the proportion of the adult population below Level 3 is 30%, 40% or 60% will have little impact on the policy options to be pursued, but will influence the magnitude of the investment needed to reduce this proportion to a target level.

Finally, the policy prescription associated with estimates of either the absolute or relative risks associated with skill-based inequalities is unlikely to change given that these risks are large by any standard. More directly, the policy options to respond to adjusted log odds of experiencing poor outcomes of 2.5 times will be no different than one that would respond to an adjusted log odds of 4 or 5 times.

A deeper analysis of the needs of countries for reliable data on literacy and numeracy distributions suggests a second set of needs to be met.

Less developed countries need more than a tool to tell them that they have a problem and to attract their fair share of aid. They also need a means of translating the policy implications of the assessment results into information that serves to transform teaching and learning.

In their native form, assessment results do not speak to educators so there is a need for a semantic layer that translates the data-driven policy messages into language that educators can understand and act upon. This act of translation involves several linked steps including the development of:

- An alternate vocabulary to describe assessment results that links to teaching and teaching content to which educators can relate
- Tools that analyse error patterns in assessment results that indicate where remedial instruction of teachers might yield rapid improvement in teaching quality and, thereby, skills acquisition
- Assessment tools that identify the life-long learning needs of individual, that document learning gain and that certify skills for employment. These tools should certify both technical skills and knowledge and the cognitive skills needed to apply those skills. The assessments of cognitive skills should report place people on the international proficiency scales to provide context and estimates of learning gain estimated through pre and post testing of learners should yield reliable estimates of point gain on the international scales. This would provide learners, training providers and funders with a way to reliably gauge the relative efficiency and effectiveness of programs.
- The development of remedial curricula that can be embedded in existing primary, secondary, tertiary and adult curricula

The development and deployment of these tools would greatly increase the return on national investments in data collection by shifting the focus from describing skill problems to fixing them. UNESCO could play an important role in facilitating the development and deployment of these tools.

The Government of Canada has invested significant resources in the development and validation of efficient and effective instructional responses to the literacy and numeracy learning needs of adults. This research has yielded significant (300% - 500%) improvements in instructional efficiency and effectiveness for a broad range of low-skilled adults (DataAngel, 2017). It is possible to raise the skills of mid-level 2 adults by an average of 40 points in as little of 15 hours of instruction. These efficiency gains open the way to much larger and more productive public investments in adult skill upgrading. Importantly for the current discussion, a key element of the new instructional recipe is the systematic pre-and post assessment of learners to adjust the instructional response to assess learning needs and to document learning gain using assessments linked to the PIAAC framework and proficiency scales.

An additional way in which the salience of the assessment results might be made more useful to educators would be to link the assessment results explicitly to the US O-Net skill taxonomy.

Currently O-Net includes several dimensions that are used to classify the cognitive demands of jobs:

- The incidence of skill use

- The frequency of skill use
- The criticality of skill use/implied mastery level

Two dimensions are currently excluded from the O-Net classification of cognitive demand of jobs:

- The complexity of skill use, as reported in the proficiency levels
- The dependence on other skills

Were UNESCO to work with the US Government to add these dimensions, the data on literacy and numeracy skill supply could be related directly to the data on the economic demand for literacy and numeracy skill implied in the distribution of employment by occupation. It would also, through Bloom's taxonomy, be directly relatable to curriculum.

3.1. The statistical implications of intended data uses

Each of the data uses enumerated above place a distinct set of statistical demands on the measures.

Meeting the latter two objectives – profiling determinants and outcomes - implies a need for the application of multivariate methods that demand relatively small sample sizes i.e. roughly 60 completed cases in each cell to be included in the analysis¹.

Meeting the former objective - of generating point estimates of average scores and numbers and proportions of the population at proficiency levels – requires higher sample sizes i.e. 100 to 400 completed cases per population subgroup for which data is needed by design. For this reason, international adult skill assessments have tended to field average samples large enough to yield completed cases for 5,000 adults.

3.2. Antecedents of PIAAC: YALS, LSUDA, IALS, ALL, ISRS, LAMP and STEP

The OECD's PIAAC assessment is by far the world's most visible international assessment of adult skills, one that yields demonstrably valid, reliable and comparable estimates of literacy and numeracy skill and for the key outcome and determinants co-variates needed by policy makers. The PIAAC program is the result of a remarkable 35-year international collaboration. PIAAC can be traced back to the Young Adult Literacy Survey (YALS) conducted in the United States in 1985. It was the first national study to employ the theory of adult reading developed by Irwin Kirsch and Peter Mosenthal and to produce literacy and numeracy proficiency estimates using 3-parameter Item response models applied to representative samples of young out of school adults. The YALS study assessed prose literacy, document literacy and quantitative literacy.

The Government of Canada subsequently conducted a national assessment, known as the Literacy Skills Used in Daily Activities Survey (LSUDA) in 1987 in two languages: English and French. LSUDA assessed prose literacy, document literacy and quantitative literacy and results were reported on the YALS scales.

The National Adult Literacy Survey (NALS) was subsequently fielded in the US on a large sample of adults. NALS assessed prose literacy, document literacy and quantitative literacy. NALS created new, more robust proficiency scales.

The International Adult Literacy Survey (IALS) was then developed and administered in several waves in a large number of countries and languages. Importantly, participating

¹ Assuming a design effect of 2

countries contributed test items that were added to the international item pool. IALS assessed prose literacy, document literacy and quantitative literacy that were reported on the NALS scales. Data is available for 25 countries. The NALS design was adopted by IALS after a systematic review of approaches to assessment that had been applied globally (Satin and Murray, 1990). Only the NALS approach was judged to satisfy the statistical demands of providing comparative results in multiple languages.

The Adult Literacy and Life Skills Survey (ALL) was developed and administered in two waves – 2003 and 2005. The ALL study retained the NALS/IALS prose literacy and document literacy scales and replaced the quantitative literacy scale with a broader numeracy scale. Problem solving was also assessed. 26 countries participated in the development of the assessment items included in the ALL assessment. The ALL assessment was also fielded in Bermuda and in the Mexican State of Nuevo Leon.

The International Survey of Reading Skills (ISRS) was developed and implemented in 2005 in Canada and the US by Statistics Canada, NCES and ETS. The ISRS study was designed to assess the component reading skills thought to underlie the emergence of fluid and automatic reading that is needed to master Level 3 and above literacy and numeracy tasks i.e. letter and number recognition, receptive vocabulary, decoding fluency and accuracy and passage fluency. The availability of these measures provided deep insight into the learning needs of Level 1 and 2 adults, a part of the NALS/IALS/ALL proficiency distribution about which little was known.

The LAMP program adapted the ALL/ISRS methods and assessed prose literacy, document literacy and numeracy, as well as a variant of the ISRS reading components, in 10 countries and 13 languages. Participating pilot countries developed additional test items to both acquaint them with the underlying framework and to add face validity to the item pools. Results were reported on the three (prose literacy, document literacy and numeracy) scales but was not linked to the IALS/ALL scales. Comparative analyses of reading components were undertaken for four countries who have conducted the main survey. In each country, the respondents who have taken reading components are classified into groups and described using latent class model, but reading components are not put on the comparable scales.

The first wave of PIAAC collection was conducted in 2011. In order to free up assessment time PIAAC combined the NALS/IALS/ALL prose literacy and document literacy scales but retained the ALL numeracy scale. PIAAC also assessed problem solving in a technology rich environment.

The PIAAC design also provides data on reading components – the skills that underlie the fluid and automatic reading that characterizes level 3 proficiency on the international literacy scales. These data only support second order comparisons. For example, they only allow a comparisons of estimates of the proportion of the target population that recognize 80% of the symbol set in a given language, not the absolute number of symbols.

The PIAAC reading component assessment was adapted from measures developed for the 2005 International Survey of Reading Skills (ISRS) jointly developed and implemented by the Canadian and US governments in collaboration with the Educational Testing Service.

Importantly, PIAAC was the first adult assessment to include test administration on a computer platform. Analysis of the resulting data suggests that the paper and pencil and computer-based collections yield comparable results even for very low-skilled adults.

The STEP program adapted the IALS/ALL/ISRS/LAMP methods and item pools for use in a group of low-middle income countries. In order to reduce the operational and financial burden of fielding a comparative assessment the STEP program the STEP sample was

limited to urban areas and the coverage of the skill assessment limited to one skill domain, literacy. The STEP assessment also included the reading components assessment to improve the information yield of the assessment in the lower regions of the literacy scale.

Collectively, the YALS/NALS/IALS/ALL/ISRS/LAMP/PIAAC/STEP programs have provided a wealth of valid, reliable, comparable and interpretable data that:

- Confirms the theories of task difficulties that underlie the items
- Provides empirical confirmation that the test is working in a stable way in heterogeneous populations within and between in a wide range of languages.
- Confirms the existence of meaningful differences in the average levels and distributions of skill both within and between countries
- Confirms that these differences in the level and distribution of skill are associated with meaningful differences in outcomes and underlying policy choices

The fact that these differences in the level and distribution of skill and outcomes have been shown to be causal and tied to broadly defined policy choices creates a moral/ethical obligation on the part of policy makers to focus their attention on fostering skill demand, in creating more and more equitably distributed skill, in improving the efficiency of markets that match skill supply and demand and in ensuring maximal skill utilization.

3.3. The RAAMA study

Of the international studies reviewed for this paper the UIL RAAMA study stands apart from IALS, ALL, PIAAC, STEP and LAMP as the items used to assess literacy skill were not selected in a way to provide systematic coverage of the characteristics that underlie the relative difficulty of tasks nor were results summarized using methods that confirm the stability, reliability and comparability of measurement. The RAAMA approach to measurement does not provide the needed cross-national comparisons of skill over time. However, the content framework developed for low literate population in literacy program may be contributed to the development of the conceptual framework.

3.4. National assessments that measure skills directly

As noted above literacy and numeracy skills have been shown to play a key role in economic development and social development. In response, several countries have developed and conducted national skill assessments.

Of the national literacy and numeracy direct assessments reviewed, several studies apply methods and measures that are conceptually close to those used in the YALS/NALS/IALS/ALL/PIAAC/LAMP/STEP assessments but none offer results that support comparison to other countries data. For example, the assessments fielded in Scotland, Kenya and Bangladesh drew heavily on IALS and ALL items.

3.5 National assessments that measure skills indirectly

A number of national assessments were reviewed that measured literacy and numeracy skills indirectly. These assessments rely on self-reports of skill, or on performance on very limited numbers of test items. Research by the author shows that these measures are unreliable i.e. they are unable to support comparisons within or between countries (Neice and Murray, 1997). The fundamental problem with self-reports is that adult perceptions of their skill levels are conditioned by their use of their skill rather than their actual skill level and at times social perception of having the skills. To make matters worse, the relationship of self-perceived skill to actual skill varies significantly among sub-populations within countries and across countries and over time. This renders these assessments of limited use for policy makers.



The fundamental problems with assessments that use small numbers of test items to determine skill levels is that they offer little discrimination along the continuum of skill and offer too little information to support any statistically defensible generalization of proficiency level. These studies also are much more susceptible to cheating.

As a result, these assessments are of little use in defining thoughtful policy responses.

4. Definitions of functional literacy and numeracy

The definition of literacy have evolved considerably over the past 40 years in response to theoretical advances that allow one to predict the relative difficulty of reading and numeracy tasks to a high level of precision.

The key insight afforded by these theoretical advances is that attributes of the text being read have very little impact on the relative difficulty of tasks, explaining only 15% of observed variance in task difficulty. Rather, the cognitive demands of the reading task explain the overwhelming majority of differences in task difficulty (see The Adult Literacy and Life Skills Survey: New frameworks for Assessment (2005) Statistics Canada and NCES).

The current predictive models upon which the PIAAC literacy explain an estimated 93% of the variance in task difficulty in the range of skill observed in OECD countries and thus provide assessment designers with the means to design assessments in which the test items systematically sample the attributes that predict their difficulty. It is this design feature that lays the foundation for the interpretation of test results as a reliable reflection of adult's reading ability. As noted above, mastery of the linguistically idiosyncratic reading components measures explain the emergence of fluid and automatic reading needed to support the mastery of level 2 reading tasks that require readers to locate one or more pieces of information in a text.

Importantly for current purposes, analysis of data from the IALS, ALL, PIAAC, STEP and LAMP assessments confirms that the models that predict task difficulty are stable across a diverse range of language and culture that including English, Arabic, Chinese, Japanese and Korean.

Although IALS, ALL, PIAAC, STEP and LAMP all use different words to describe what has been measured, the actual way in which literacy and numeracy are measured in all four studies is identical.

The PIAAC framework begins by **defining literacy** as "understanding, evaluating, using and engaging with written texts to participate in society, to achieve one's goals and to develop one's knowledge and potential."² This definition implies far more than just reading the words of the text. It includes an emphasis on how the information gathered from this encounter with written materials are used and influence one's thinking. It is also very inclusive of all kinds of text – not just words on a page, but also visual displays, graphics, as well as materials that are presented through a digital medium.

The choice of **texts** is wide ranging and is categorized according to six **characteristics**:

- medium – print and digital
- format – continuous and non-continuous
- type – rhetorical stance (i.e., description, narration, exposition, argumentation, instruction, and records)
- physical layout – type of matrix organization
- features unique to digital text – hypertext, interactive, other navigation features (use of scroll bar, utilization of mouse).

There are three **task aspects** that are specified in the PIAAC framework. These three broad cognitive strategies are designed to achieve a full understanding of the texts:

² Ibid (p. 20)

- **Access and identify** tasks require the reader to locate information in a text. These items include both locating a single piece of information and cycling through the text when more than one piece of information is necessary.
- **Integrate and interpret** tasks require the reader to understand the relationship between parts of the text (i.e., problem/solution, cause/effect, etc.).
- **Evaluate and reflect** tasks call on the reader to draw on information, understandings, and knowledge that is external to the text. These tasks include such things as determining the relevance, credibility, etc. of a text. It also includes tasks that look at the register, structure of the text as well as the types of evidence that is provided.

Systematic sampling of this design matrix allows the PIAAC proficiency scores to be interpreted as reliable indicators of general proficiency.

Current approaches to assessment, such as the PIAAC design, also embody a second scientific advance, specifically the development of a set of statistical methods that extract a reliable proficiency score out of statistically error laden test item data. Basically these methods allow one to reliably order test items by their relative difficulty and to place individuals on the same scale. Having items and people on a common scale provides deep meaning to the proficiency estimates, allows one to apply a mastery level to the proficiency estimates and provides empirical confirmation that the test is working in the same way in all populations and subpopulations. This latter characteristic is fundamental to comparing the results across languages, populations, sub-populations within countries and over time.

Thought of in statistical terms, the approach to measurement and the derivation of proficiency scores must be capable of generating comparable results over heterogeneous populations, over time and from diverse samples of test takers. Importantly, the statistical methods applied must yield definitive evidence that the test is proving stable measures within and between populations and over time.

This theory and the associated statistical methods have been successfully applied in virtually every international comparative assessment of literacy, including PIRLS, PISA, IALS, ALL, PIAAC, STEP and LAMP. These studies display a coherent approach that links definition to theory to measurement.

A review of national and international direct assessments that were reviewed for this paper reveal several common themes.

First, most embody a notion of literacy as a continuum, rather than a dichotomy

Second, most include a notion of understanding and application of skills and of coping with the social and economic demands faced by individuals.

Unfortunately, most of the studies fail to apply tests that provide good coverage of either the variables that underlie task difficulty, nor of a balanced range of life contexts.

Moreover, only PIAAC applies methods that allow results to be compared over time and across countries at very different levels of educational development.

5. The way forward: Recommendations for definitions and measurement

Given the size of the investment made in developing and validating the methods deployed in PIAAC, the wealth of useful data that already exists, and the low likelihood that an investment would improve on the scientific rigour of the approach to measurement an obvious option would be to recommend that all countries administer the PIAAC assessment since it provides demonstrably comparable estimates of literacy skill average scores and distributions by proficiency level as well as extensive data on skill determinants and outcomes. The PIAAC measures are:

- Valid, in the sense that they are based on proven theory
- Reliable, in the sense that the empirically observed item difficulties match the theoretically predicted item difficulties closely
- Comparable, in the sense that the statistical methods applied to derive scores provide empirical confirmation that the test is working in a stable fashion within and between countries and over time.
- Interpretable, in the sense that proficiency estimates are linked to material outcomes and provides insight to the design of remedial skill upgrading programs
- Affordable, in the sense that the considerable design costs and implementation costs are amortized over a large number of participating countries.
- Manageable, in the sense that the probability of experiencing catastrophic errors in implementation is within acceptable limits

The PIAAC program also offers an implementation process that both affords an opportunity to have input, through the Board of Participating Countries, into decision -making and that includes quality assurance measures that reduce the risk of a participating country experiencing a catastrophic error in implementation.

This latter attribute is important for non-OECD countries thinking about participating in PIAAC because experience suggests that the PIAAC methodology is sufficiently operationally, technically and financially demanding to overwhelm even the most developed of countries. For example, France, Italy, Russia and Indonesia have all experienced serious technical failures despite having access to considerable technical expertise.

This fact implies a need for less-economically and technically developed countries to consider measurement options that impose lower financial, operational and technical demands whilst still satisfying country's information needs.

It is worth noting several more subtle implications of applying the IALS, ALL, PIAAC, STEP and LAMP methods.

The first implication is that the underlying conception of literacy shifted from a dichotomy of literate/illiterate to one in which people were classified into groups based on the probability of them being able to master reading tasks at a given level of difficulty, where the desired level of mastery is chosen to reflect the point at which classification errors are minimized and the cost of getting a task wrong are within acceptable limits.

The second implication is that the theory that allows for skill to be measured along this continuum is that it allows one to identify thresholds that define proficiency levels that reflect shifts in the underlying cognitive demands of reading tasks. This relationship opens the way to more efficient and effective approaches to instruction. These same levels have

also been shown to be tightly tied to differences in individual educational, health, labour market and social outcomes, so provide a means to focus policy on those most in need.

The third implication is that the statistical methods applied to summarize proficiency allow results to be safely compared within and between populations and, most importantly, over time. Data on skill supply only makes sense when compared to estimates of skill demand for social, educational, democratic and economic ends and indices of skill usage. The fundamental questions for policy makers are whether skill supply is able to keep up with changes in skill demand and what needs to be done if skill shortages reach economically and socially unacceptable levels. The advance of technology is likely to rapid increases in the levels of skill demanded by employers over the medium term and will amplify current levels of skill-based inequality in outcomes.

None of the other studies reviewed provide a means to compare results over time or across countries – both key requirements for monitoring Sustainable Development Goal 4.6.

5.1. The statistical demands of the Sustainable Development Goal target 4.6 and the associated indicator

The Sustainable Development Goal target 4.6: states that “By 2030, ensure that all youth and a substantial proportion of adults, both men and women, achieve literacy and numeracy.” The global indicator for the Target 4.6, and the only indicator for this target directly related to the measurement of learning outcomes, is the indicator 4.6.1, which states: the percentage of the population in a given age group achieving at least a fixed level of proficiency in functional (a) literacy and (b) numeracy skills. The target age group for this indicator is the population of 15 years and older.

Translated into statistical terms the indicator for Target 4.6 implies a need for:

- Separate measures of literacy and numeracy
- Measures that are statistically representative of the adult population
- Measures that capture the full range of skills possessed by the adult population
- Measures that can be safely compared, at a point in time and over time
- Measures that are sufficiently precise to detect economically and socially meaningful change in skills over the reference period.

The indicator specification also includes several subjective elements that require definition including:

- What constitutes a “substantial” proportion of adults and youth?
- What is the underlying meaning of “achieve” literacy and numeracy?
- What is the definition of “functional” relative to literacy or numeracy?

Defining “substantial”

In this context, it is difficult to determine what the term “substantial” should mean. The main issue lies with the fact that the definition of substantial will depend on the level of economic and educational development and national economic and social objectives. For example, in the economically and educationally advanced countries of the OECD substantial in 1950 would have implied a need for 60% - 75% percent of the youth cohorts graduating from secondary school and 15% having the skills to go on to post-secondary study, targets that imply two levels of literacy and numeracy. Now the norm in these same countries demand 95%+ secondary graduation rates and up to 85% of youth cohorts going on to some form of post-secondary study. In this sense “substantial” will vary from country

to country in ways that reflect their underlying social, educational and economic objectives.

Defining "achieve"

The term "achieve" carries many meanings but in this context it is taken to imply a level of mastery.

The OECD PISA study chose to adopt a probabilistic mastery standard commonly used in school settings i.e. individuals getting 62.5 % of items at a given proficiency level are placed at that level. In practical terms, this standard is quite low as individuals can be deemed competent even if they get 37% of items at the target proficiency level incorrect. Ironically, the cause of adult literacy and numeracy shortages may lie in the low standards imposed by the primary and secondary education systems.

The IALS and ALL study chose to adopt a more demanding mastery standard, one borrowed from the trades in North America. To be judged to be proficient tradespeople needed to demonstrate mastery of 80% of the required content, including any tasks requiring reading and numeracy.

PIAAC chose to relax this mastery standard to that applied in the OECD PISA study. While this makes the proficiency estimates from PISA and PIAAC more aligned, this does a disservice to the data, in the sense that very few employers would tolerate employees that got 37% of their reading-based decisions wrong. In addition recent analysis by the author suggests that the low levels of mastery demanded by the education system in Canada actually induce employers to lower the technical and cognitive demands of their jobs to cope with high proportions of workers with unreliable skill levels. This would not be a problem were it not for the facts that this behavior is associated with a significant loss of output per hour worked and massive skill loss on the part of the workforce that initially had the required, skill level.

In reality, however, both of these mastery standards do a disservice to the underlying data if one considers the literacy demands of work. An empirical analysis by the author identified the mastery level that yielded the most reliable classification of proficiency levels in different occupations. The analysis suggests that the optimal mastery level – i.e. the one that yields the cleanest classification of individuals by level – actually varies by occupation. For occupations with high knowledge and skill intensity where the costs of error are high – such as neurosurgeons and other medical specialties – the optimal mastery standard was 95%. For occupations at the other end of the spectrum, where the costs of error were relatively low – such as labourers in agriculture - the optimal mastery level is 5%. Workers in these latter occupations could afford to get 95% of their reading tasks incorrect and still be judged competent.

Since no country collects systematic information on the costs of literacy and numeracy errors, there is little choice but to adopt the prevailing PIAAC mastery standard as a prerequisite to comparison to PIAAC proficiency estimates. Nonetheless, care will have to be taken in making definitive statements about the adequacy or functionality of any given national proficiency distribution.

Defining proficiency levels

In the NALS/IALS/ALL/PIAAC/LAMP/STEP framework proficiency levels are not defined by the imposition of arbitrary cut points on the scale but are defined by points where the underlying cognitive processes shift. Individuals are then placed at a given proficiency level probabilistically by meeting or exceeding a set proportion of items at that level correct. Provided that the fit between predicted and empirically observed item difficulty are in close agreement then placement can be interpreted as a general indication of

proficiency. This approach works well in the upper regions of the scale where cognitive strategies explain virtually all of the observed difficulty of items. The available data suggests that these strategies are universal.

Placement in a level in the lower range of the scale is more complex as the predictors of task/item difficulty explain much less of task difficulty. In this region of the scale analysis of reading components data for multiple language suggests that performance depends on mastery of the reading components. Adults who have yet to master the reading components are still in the process of learning to read and must devote most of their cognitive resources to the act of reading. In contrast, adults that master all of the components are, by definition, fluid and automatic readers who can devote the bulk of their cognitive resources to building meaning. Since the complexity of these component tasks varies by language, the component tasks can only be compared at the second level e.g. what proportion of the population can recognize 80% or more of the symbol set used to represent the language. Analysis of data for Canada identifies six specific groups of learners, each with a unique pattern of strength and weakness in their reading components and each of which demands a unique instructional response. In other languages analysis of the reading component data will reveal different groups.

In this sense the lower regions of the NALS/IALS/ALL/LAMP/PIAAC/STEP scales are not strictly uni-dimensional.

It is recommended that the PIAAC proficiency levels be adopted but that the reading components data for each country be analyzed to reveal groups of learners sharing common patterns of strength and weakness in the target languages. The numbers and proportions of adults in each such group should also be reported annually.

Defining "functional"

A second related implication of the NALS/IALS/ALL/PIAAC approach to measurement is a shift from an arbitrary definition of what constitutes being functionally and fully literate. In the past, whether someone was functionally or fully literate depended on the imposition of arbitrary score cut points on an arbitrary proficiency scale. For example, the UNESCO definition of literacy was based on years of schooling, with adults having completed 4 years of education as "literate" and adults having completed 9 years of education being classified as "fully literate". Analysis of data from the 1994 IALS study revealed that this indirect measure resulted in massive misclassification in both directions even in the economically-advanced countries of the OECD.

In the current generations of comparative assessments, functionality is defined in a variety of ways including relative to the reading demands of their jobs, to the level that reduces the probability of experiencing poor outcomes and allow people to achieve their goals.

A third, even more subtle, implication of the current approaches involves a shift from a sociologic view of literacy to an economic view of literacy. The underlying goal of the sociological view is simply to describe what reading tasks people do in their daily life and, by extension, their ability to cope with the familiar. In this view everyone has their own literacy that is neither better nor worse than someone else's literacy. The sociological view ignores the fact that literacy is, by definition, a mechanism for the transmission of information among people. To serve this purpose, information theory suggests that the structures upon which literacy is based need to be regular and shared.

In the economic view of literacy the underlying goal is to judge people's ability to cope with unfamiliar tasks that they encounter in daily life. The goal is not to access the familiar but to acquire new knowledge or information. As noted above, the scientific advance that underlies the comparative assessment of literacy and numeracy is the development of theory that reveals the characteristics of task and text that underlie the relative difficulty

of reading and numeracy tasks. The implicit assumption in the economic view of literacy is that there are economic consequences that accrue to the individual, the family, the community, to social institutions and to the society and to the economy that are associated with being unable to cope with unfamiliar reading tasks at each proficiency level.

In the economic view, literacy is best thought of as tool that allows individuals to acquire new technical skills and knowledge efficiently and effectively and to apply these skills and knowledge efficiently and effectively once acquired.

The economic view has also be useful to describe the systems that generate literacy skill, the systems that utilize skills once acquired and the markets that match the demands of jobs with worker skills. These market models of skill have revealed significant inefficiencies in all three of these systems, information that is crucial to the development of efficient and effective policy responses.

It is important to remember, however, that literacy is not the only way to acquire and apply technical skills and knowledge. Homo sapiens have a long history of observation and transmitting skills and knowledge orally. This fact complicates the measurement of the skill levels of adults with level 1 and 2 literacy skills. Tasks at this level only require readers to locate single or multiple pieces of information. Adults and youth can learn to cope with this level of tasks without actually knowing how to read. Yet, adults with functional literacy skills have been shown to be more efficient learners, more economically productive, healthier, more socially engaged and less socially and economically dependent.

Judging whether someone is functionally or fully literate is, by definition, a relative concept that involves the comparison of the level of reading demands faced by individuals and the skill level they have. This comparison allows individuals to be classified as being in literacy skill shortage, balance or surplus and for an analysis of the impact that each has on economic and social outcomes observed at various levels.

Increasingly life requires individuals to read, understand and apply information derived from print. Individuals who have the skill levels needed to cope reliably with familiar demands they encounter gain agency and independence. Those who lack the level of mastery needed to cope with unfamiliar reading tasks at a given level of difficulty risk making the wrong decision or of relying on others. In the former case the cost of error may be high, in the latter dependence creates a power relationship in which one cannot assume that interests are perfectly aligned.

So the first level of functionality is determined by whether individuals have the skill to meet the current reading demands they face in their daily life. Important to note that the proficiency level demanded will vary from individual to individual and some individuals manage to live perfectly satisfying and productive lives without any reading whatsoever.

The second level of functionality is related to literacy as a tool for individuals to deal with unfamiliar reading demands, either proactively or reactively.

Reactive uses of literacy involve individuals having the proficiency level to cope with reading demands that are imposed by external forces – getting a disease and having to understand and apply dosage and contraindications on medications, learning to use a new piece of machinery safely, etc. The available evidence suggests that Individuals who do not have the required level of proficiency bear a significant burden judged in terms of poorer outcomes. They are less employable, work less, earn less, are in poor health, more at risk of experiencing a workplace illness or accident and are less socially engaged.

Proactive uses of literacy involve applying one's literacy skill to achieve one's goals – the definition of literacy that is embedded in all of the international comparative assessments of adult literacy, including the OECD's 2012/2014 PIAAC assessment cycles.

In this version of functionality, individuals need a given level of reading proficiency to give them a reasonable probability of realizing their goals. For example, college programs require a minimum of Level 3 reading proficiency to get full value out of the experience and to have a reasonable probability of persisting to the point of graduation. Individuals without this proficiency level realize demonstrably poorer educational outcomes.

Thought of in economics terms, in this version of functionality individuals need a means to:

- Know their literacy proficiency level,
- Understand the proficiency level needed to have a reasonable probability of realizing their goals,
- Understand what type and level of investment would be needed for them to acquire the needed level of proficiency,
- Identify and select a training provider to fill literacy skill gaps efficiently and effectively

Unfortunately, research suggests that few individuals in need of skill upgrading have access to this information, a situation that leads us to conclude that the market for literacy skill upgrading is in failure in most, if not all, countries (CLLN, 2015). To make matters worse, the available evidence suggests that automation is likely to reduce the number of jobs that only require the routine application of procedural knowledge that is supported by Level 2 literacy and increase the demand for workers capable of fluid problem solving that requires Level 3, 4 and 5 level skill. So the proficiency level that is needed to obtain and keep well paying and stable work is rising rapidly. It is likely that large swaths of the world's population have yet to apprehend this shift or what it implies for their need for literacy and numeracy skill upgrading.

The third level of functionality moves up from the individual to the level of social organizations - for example firms, education and training providers, hospitals. At this level, there are two sides to functionality. One side is the level of literacy skill assumed by these institutional providers of goods and services. The second side of this level of functionality is the literacy proficiency level needed possessed by the clients being served by each institution. Inevitably, the literacy proficiency assumed by the providers will be above the proficiency level possessed by some proportion of the clientele.

These misfits between skill supply and demand impose costs on the institutions and on the individuals. For example, firms whose workers lack the literacy skill needed to perform at the production frontier will be less productive and profitable than their competitors who have smaller literacy proficiency gaps. Similarly, providers of health services that are largely publically funded will face higher levels of demands for their services and higher costs of treatment than providers where skill supply and demand are in balance. In these cases, both the public and private rates of return will be lower than they might otherwise be. This creates a public interest in policy measures that serve to improve the fit between literacy demand and supply. These measures include both measures to reduce the level of literacy proficiency demanded, for example plain language initiatives, or measures that reduce the skill gap by increasing the supply of literacy skill, or of measures to improve the efficiency of the markets that match skill supply and demand. Thus, by way of example, the functional level of literacy needed to maximize private and public returns on investments in the provision of public goods and services might be higher or lower than that level demanded to maximize individual level returns on literacy acquisition and application.

The fourth and final level of functionality moves the standard up to the societal level. A fundamental role of every government is to seek to improve the welfare of their citizens and to set goals for their improvement be it income, health, social welfare or social equity.

Research clearly shows that average literacy skill proficiency level and the proportion of adults with low literacy skills are the single most important determinants of differences among countries in long term rates of GDP and labour productivity growth.

Aggregate differences in the skill levels of employees of social institutions such as firms, hospital and schools, and in the skill levels of these institutions clientele, have been shown to influence both public and private rates of return on investment.

Individual skill differences have also been shown to explain most of the social inequality observed in a broad range labour market, health, educational and social outcomes. The net result is that low skilled adults bear a disproportionate share of poor outcomes. By way of example, the probabilities of Canadian adults with Level 1 and 2 literacy being in fair or poor health, of experiencing a spell of unemployment, of being poor and of being socially disengaged are roughly 2.5 times higher than their more skilled peers even after adjusting for the impact of a broad range of related characteristics such as education, age, gender, language, ethnicity, occupation and immigrant status. These probabilities are amplified in less economically developed and educated countries. This level of skill-based inequality is unfair and places a moral obligation on governments to increase the skill levels of low skilled adults, particularly because these relationships have been shown to be causal (i.e. low literacy skill causes poor health, labour market, educational and social outcomes) and the instructional approaches exist to raise adult skill level rapidly and at low cost (SRDC, 2015; AWES, 2017).

By way of summary, functionality can be only be defined in relative terms based on whose interests are being served. More directly, the level of literacy needed for a individual to cope with the reading demands that they confront in their daily lives will differ from the level needed to realize their goals, from the level needed to for social institutions to be efficient and effective and for societies to realize their social and economic objectives.

This implies a need for each country to establish their own definition of what level constitutes the functional level(s), ones that reflect their tolerance for literacy skill-based inequality in individual outcomes, their targets for the performance of key social institutions, including firms and educational institutions, and their social and macro economic goals. Because some of the impacts of literacy skill are collective – either influencing the public’s return on tax investments or the overall levels of macro-economic performance, social progress and population health governments need to implement this process.

5.2. The definition of numeracy

To this point, this paper has focused on defining and describing the approaches to measuring literacy. This section focuses on numeracy and numerate behavior.

The 2011 PIAAC assessment framework defines numeracy as follow:

Numeracy is the ability to access, use, interpret, and communicate mathematical information and ideas, in order to engage in and manage the mathematical demands of a range of situations in adult life.

The PIAAC framework goes on to define numerate behaviour:

Numerate behavior involves managing a situation or solving a problem in a real context, by responding to mathematical content/information/ideas represented in multiple ways.

The PIAAC numeracy definition mirrors the literacy definition in that it implies the acquisition and application of a set of skills in response demands placed upon them.

The PIAAC approach to measuring numeracy also mirrors that used to assess literacy i.e. authentic tasks are developed that systematically sample the characteristics of text and task that underlie, or predict, relative task difficulty (and, by extension, proficiency).

The PIAAC approach to measurement also employs the same statistical methods to summarize proficiency and to provide empirical confirmation that the assessment items are performing in a psychometrically stable way both within and between countries and over time.

Thus, as for literacy, the implementation of the YALS, NALS, IALS, ALL, ISRS, LAMP and PIAAC studies offer proof that the measures are valid, reliable, comparable and interpretable proficiency estimates. It is reasonable to recommend that countries consider participating in the next round of PIAAC implementation.

It is worth noting three things about the PIAAC notion of numeracy in passing.

First, that the PIAAC numeracy proficiency measures depend, and end up being highly correlated with, the literacy measures, simply because test takers must read and understand both the nature of the question being asked and the type of response expected.

Second, despite the high level of correlation between the two measures, numeracy skill does have an independent and significant impact on outcomes. Numeracy is key to the application of the scientific method and to critical thinking. In addition, the underlying theory defines a unique set of variables that define both acquisition of numeracy skill and their application.

Third, analysis of health literacy undertaken by the author reveals that numeracy skills dominate the relative difficulty of more difficult health literacy tasks. This finding suggests that numeracy skill provides adults with a more efficient tool for solving complex problems. This fact implies that numeracy skill will become increasingly important to individual and collective outcomes as technology takes over Level 1 and 2 literacy and numeracy tasks.

6. Availability of the range of items from diverse adult literacy assessments that could be used to represent the measure of adult functional life skills and competencies and could be used to benchmark minimum proficiency level

One objective of this paper is to evaluate if the current item pools contain sufficient items to measure adult skill levels and to benchmark minimum proficiency levels.

A simple reading of the previous chapter of this report suggests that the criteria that should be applied should include:

- Whether the stimuli and associated tasks are authentic
- Whether the stimuli and associated tasks offer sufficient coverage of the underlying variables that determine the difficulty of tasks across the full range of proficiency
- Whether the item pool is equally unfamiliar to low skilled respondents in the target countries as one of the defining features of high skilled adults is that they can cope with unfamiliar texts and tasks. This criterion could be met simply by offering participating countries an opportunity to develop new items that reflect the framework. While it is highly unlikely that these items would influence the measurement, these items would add face validity and provide first hand insight into the literacy framework and approach to measurement. The matter is, however, significantly more complex than it first appears. Previous generations of adult literacy skill assessment provided very little measurement below 250 on the 500-point proficiency scale, a fact that greatly limited the utility of the assessment data for policy makers interested in fixing the skill issues identified by the overall assessment even in highly literate and educated populations. The information yield of these assessments, for both individuals and countries scoring in the lower regions of the scale, was quite limited. As described below, the Canadian and American governments took steps to correct this weakness through the development and administration of a battery of tests that tested individual's mastery of the skills that underlie the emergence of fluid and automatic reading that is a prerequisite for coping with unfamiliar Level 3 tasks.

As noted above the items administered in current generation of assessments are designed to systematically sample the attributes of task and text that underlie the relative difficulty of reading tasks.

Conventionally adult proficiency scores have been placed on a 500-point scale. To avoid bias in the computed average scores that form the first level of comparison, the items need to cover the full range of skill in the observed populations with limited ceiling and floor effects i.e. negligible proportions of respondents who get all the test items right or all the test items wrong.

The item pools fielded in the initial IALS and ALL assessments were selected to provide full coverage of the range of literacy skill observed in the participating countries who were predominantly OECD member states. The selected items were, however, developed by a much larger group of countries and in a wide range of languages (i.e. 26 countries and 16 languages) in an effort to improve the representation of items from non-OECD countries. Implementation of the LAMP and STEP programs have extended the linguistic and cultural coverage of the item pools.

Items were also selected to provide balanced coverage of a range of life contexts and materials so that no individual or population would be systematically disadvantaged

because of their unfamiliarity with a given type of content. Importantly, all selected items were based on authentic texts and real life questions.

Notwithstanding the IALS and ALL satisfied the technical requirement to provide coverage of the full proficiency distribution analysis of the data suggested a need to administer additional measures in the lower regions of the proficiency scale to provide more insight into the learning needs of adults scoring below 275 on the international scales. Effectively, any individual scoring below this proficiency level was placed there deductively because of their inability to demonstrate mastery of Level 3 tasks. Thus, additional insight was needed to inform the design of remedial instructional programs. In response to these needs Statistics Canada and the US National Centre for Education Statistics fielded the International Survey of Reading Skills (ISRS). The ISRS was the first large-scale assessment of reading components – the skills that readers need to master to be fluid and automatic readers who can apply all their cognitive resources to applying what they have read.

As planned, the reading components data filled an important gap in the IALS/ALL proficiency scale. The conversion of the IALS/ALL/PIAAC proficiency scales to true interval scales allow the results of the two sets of measures to be reported on a common, single scale.

Analysis of the ISRS data confirmed that the majority of adults in the lower regions of the scale were still demonstrably still in the process of learning to read in the sense that they had yet to master one or more of the component skills – weak letter recognition, weak decoding fluency and accuracy, inadequate receptive vocabulary scores.

A deeper analysis of the ISRS data provided a nuanced profile of the shared learning needs of different groups of adults with low skills, each of which implied a distinct instructional response. Prosaically, the ISRS data provided the information needed by policy makers to define the nature and costs of remedial policy responses to the problem of low skills identified in the main assessment. More specifically these data allowed policy makers to identify:

- The size of the skill problem
- Groups of learners sharing common learning needs and distinct instructional responses
- The numbers of learners in each group and the cost of remedial instructional needed to upgrade their skills
- The benefits that would accrue to higher skill levels and the implied rates of return on investment in each group

This innovation, subsequently adopted by the PIAAC, LAMP and STEP assessments, also transformed the proficiency scale into a true interval scale, one in which zero represents a complete absence of literacy i.e. someone who is unable to recognize a single letter of the alphabet/symbol set in a language. This transformation also applies to the numeracy scale as the definition of numeracy includes a need for adults to read and understand the problem to be solved.

This transformation has important implications for assessment in less literate populations as it effectively supports comparison of skills across the entire proficiency distribution. It also allows scores to be aggregated to yield estimates of the total literacy skill supply and, comparison to literacy skill measures derived from the occupational distribution of employment, estimates of occupational literacy skill shortages. These aggregate estimates have proven very useful for government policy makers in focusing their skill upgrading investments.



Experience suggests a need for participating countries to acquire a deep understanding of the framework and how skill is measured. Such knowledge is best acquired through actually developing and validating assessment items that reflect the framework.

Experience also suggests a need for countries and test takers to see themselves in the item pool. The development and validation of a block of national items with high face validity would help assure some level of familiarity. While the addition of these idiosyncratic national items will have little impact on the reliability and precision of the proficiency estimates, it would improve response rates and this reduce non-response bias in the estimates.

7. Identification of measurement options

This chapter of the report offers recommendations of a pragmatic and cost-effective implementation method, and tools, to measure functional literacy and numeracy in a diverse world,

Personal experience suggests that, as useful and important population-level literacy skill assessments are to policy makers, the act of running a traditional assessment like PIAAC imposes a financial, technical and operational burden that many developing countries are unable to shoulder. Put differently, even with the best of intentions, many countries will be unable to meet the strict conditions set for reporting their results in sanctioned international publications. This suggests there are pressing need to reduce the technical, financial and operational demand of whatever assessment is undertaken. Importantly, however, these reductions need to maintain the comparability of the results both within and between countries.

Readers should note that the former comparisons – of population subgroups within counties – is far more demanding statistically than comparisons of differences in skill between counties. The variance in skill in the former is far larger than the variance observed between countries. So the fundamental design constraints are to capture the within country variance in literacy skill while maintaining the cross-country comparability of results.

7.1. Alternate Approaches to measurement

As noted above, current approaches to assessment impose a significant technical, operational and financial burden on participating countries. These burdens translate into a significant risk that the results will fail to be publishable.

The following sections explore the ways in which the current PIAAC approach might be adapted to reduce the operational, technical and financial burden of assessment without compromising the utility and comparability of the proficiency estimates. Four options are outlined below:

- Limiting the geographic coverage of the assessment
- Reducing the number of skill domains assessed
- Adopting a fully-adaptive web-based assessment
- Using purposive sampling and synthetic estimation to generate proficiency skill distributions

7.1.1 Reducing the number of domains assessed and geographic coverage of the samples

The first two ways in which one can reduce the operational and financial burden associated with the conduct of a national assessment is to reduce the number of skill domains assessed and/or the geographic coverage of the sample.

The World Bank's STEP program was adjusted in these ways to reduce the financial, response and operational burden of implementing PIAAC while maintaining a link to the international proficiency scales. These reductions are realized by restricting sample coverage to urban areas and by reducing the number of skill domains assessed from three to one. While the resulting data are still useful for policy makers the fact that the results of the STEP assessment are not representative of the entire adult population means that they do not fully satisfy the needs to monitor progress toward SDG 4.

Restricting the coverage to urban areas reduces the costs and operational burden of the STEP assessment but ignores the fact that literacy deficits are likely to be far higher in rural areas in most countries.

Reducing the coverage to a single literacy domain does not impair the utility of the assessment results much because the available evidence suggests that numeracy proficiency is highly-enough correlated with literacy proficiency that numeracy skill levels may be inferred.

The STEP study also included a variant of the ISRS/LAMP assessments of reading components, administered to all respondents. This inclusion ensured good coverage of the lower regions of the literacy proficiency scale where the majority of the adult population in participating countries perform.

The level of adaptivity in the STEP literacy test is, however, limited to two stages, a feature that limits the amount of information available for adults and countries that fall outside the average level of skill assumed in the PIAAC item pool. A deeper level of item adaptivity in the design would provide results that are equally informative for all participating countries while reducing average test durations, operational burden and costs significantly.

7.1.2 Reducing the response burden and improving coverage through fully adaptive computer-based testing

One of the weaknesses in the assessment designs fielded to date is related to the fact that they incorporate only limited adaptivity in the test design, a constraint imposed by the paper and pencil method of data collection employed. The limited amount of adaptivity means that countries whose skill distribution falls in the tails of the skill distribution assumed in the selected PIAAC item pool get very little information about their skill distributions.

Thus, a second way in which the information yield might be improved while reducing test durations and sample sizes would be to adopt computer-based testing that allows much more adaptivity. Tests with more levels of adaptivity ensure that results for each individual assessed will be equally reliable and precise, something that ensures that data for population subgroups and international comparisons are equally reliable and precise.

An example of a fully adaptive test is Bow Valley College's TOWES-Prime assessments of prose literacy, document literacy, numeracy and reading components. The development and validation of TOWES-Prime was funded by the Government of Canada to provide Canadian educators with a suite of low-cost, low-burden, multi-purpose but reliable assessment tools in a range of languages. The TOWES-Prime prose literacy, document literacy and numeracy assessments each include 83 testlets, a sufficient number to provide full and balanced coverage of the entire proficiency distributions. The TOWES-Prime assessment suite tool also has several advantages, including:

- The tools yield test results in real time, something that reduces the operational and technical burden significantly while providing data to policy makers for immediate use
- The tools are available at four levels of precision to serve the full range of purposes, something that bridges the rhetorical gap between policy makers and educators
- The tools may be configured in any combination from single domain to three domain tests
- The tools are already available in 12 languages/language variants
- License fees per test are usually below the cost of printing, shipping, scoring and storing test booklets

- The assessment suite includes a test of key reading components including letter and number recognition, receptive vocabulary and decoding fluency and accuracy. This test provides full coverage of the lower regions of the PIAAC literacy proficiency scale i.e. below Level 3 where the PIAAC literacy scale offers little coverage.

Validation of the TOWES-Prime tools confirms that the systems reliance on computer delivery yields results that are directly comparable to paper and pencil testing i.e. there is no evidence of a mode effect caused by a lack of digital literacy in low skilled respondents. This said, developments in tablet technology hold the promise of approaches to assessment that greatly reduce the dissonance associated with the use of input technologies that may be unfamiliar. For example, the use of voice to provide instructions, and the use of voice and touch to capture responses, will eliminate mode effects.

Currently TOWES-Prime results are reported on the original IALS/ALL literacy proficiency scales but could easily be presented on the combined PIAAC/STEP proficiency scale. The TOWES-Prime numeracy results are presented on the PIAAC numeracy scale. Linkage to the IALS/ALL/PIAAC scales is achieved through common items equating.

The TOWES-Prime assessments have been successfully administered throughout the country in Romania.

The EU-OECD's PIAAC on-line assessment offers many of the same advantages but with significantly less adaptivity – so does not provide as balanced coverage across the full range of proficiency likely to be observed in the world's less developed economies.

Both TOWES-Prime and PIAAC on-line require stable internet connection with sufficient bandwidth to support testing. Piloting Romania and in several Caribbean countries suggests that hot spots tethered through a 3G, 4G or LTE cell network offer the needed level of connectivity. Consideration might be given to adapting these tools for delivery on hardware platforms that are less reliant on stable connectivity.

Administration of both the TOWES-Prime and the PIAAC on-line assessments would yield real time proficiency estimates at a much lower financial, technical and operational burden than the STEP assessment. New Zealand's web-based adaptive national assessment uses similar methods but does not provide for international comparisons.

Adopting the TOWES-Prime assessment platform and item pool would yield significant improvements in the reliability and precision of individual and aggregate proficiency estimates and concomitant reductions in the technical, operational and financial burden associated with fielding a national assessment.

7.1.3 Reducing the sample size through purposive sampling and synthetic estimation

A fourth way in which the operational, financial and technical burden of the IALS, ALL, PIAAC, STEP and LAMP approach to assessment might be reduced is through the application of purposive sampling and synthetic estimation.

Purposive sampling of the adult population allows one to select relatively small samples of adults with specific known characteristics that capture a significant amount of the expected variance in proficiency.

In such designs, rather than using a representative sample to generate point estimates of the number of adults at each proficiency level, the underlying statistical goal is to estimate the probability of an adult with a given set of characteristics falling at each proficiency level. These probabilities are then applied to a secondary dataset that carries the same

variables – ideally the most recent Census of Population that has virtually no sampling errors. Multiple multivariate imputation is used to put scores on each individual record on the target file and scores aggregated to yield synthetic estimates of average scores and distributions by proficiency level that are sufficiently reliable for policy.

For example, in most countries one expects to see significant variation in proficiency by education level, by urban rural and for men and women. Selecting a sample of roughly 100 adults with known characteristics for each cell in this design matrix – gender (2), education level (3) and urban/rural (2) - yields a total sample of 1200 cases that is enough to generate reliable estimates of differences in average scores and in distributions by proficiency level overall and among each cell in the design matrix.

Provided that the design co-variates capture a sufficient amount of the total variance in proficiency scores, the approach yields estimates that are good enough for informing national policy and that may safely compared.

A variant of this approach would be to select a purposive sub-sample from an existing survey such as DHS or MICS, administer a computer-based adaptive assessment to the selected respondents that includes a shortened PIAAC background questionnaire that excludes the demographic characteristics already collected by MICS or DHS. The observed skill-background covariance structure would then be used to impute scores onto all individuals on the target base file. This approach would reduce the financial, operational and technical burden associated with the conduct of a stand-alone paper and pencil assessment.

The methods used to generate synthetic estimates are well established (DataAngel, 2015). For example Canada, the US and the Netherlands have all successfully applied these methods to improve the utility of national skills assessment data by using the co-variances between assessed skills and background variables to produce two sets of products:

- Projections of literacy and numeracy skill for small geographic areas
- Projections of literacy and numeracy skill for future periods

Examples of each of these is described below.

Projections of literacy and numeracy skill for small geographic areas

In Canada the co-variances between assessed skills and a set of background variables that were available on both the PIAAC background questionnaire and the quinquennial Census of Population were used to drive a multiple imputation that placed literacy, numeracy and problem solving scores on each individual Census record. The individual records on the Census file were then used to produce skill profiles for a range of geographies from the neighborhood up and for a range of demographic characteristics. Provided that there were a minimum of 500 people in the target geography or group the estimates were sufficiently reliable to support needs assessment and funding allocation decisions. The utility and reliability of the skill estimates is directly tied to the proportion of the variance in skill explained by the co-variates and their stability over time and the fact that the Census has very little sampling error on key estimates, so errors are dominated by imputation errors that are small. Fortunately the key variables used in the multiple imputation – age, gender, mother tongue, education level, employment status, occupation - are measured quite reliably.

The best way to think about these estimates is to ask the question “Once one has controlled for a broad array of variables that capture most of the variation in literacy scores is there any reason to believe that the profile of skill in a given area would be different?” The answer to the question is “Yes, perhaps but not often enough to matter for policy”. In

practice, these demographically corrected literacy and numeracy estimates that capture most local variation in skill are far superior to applying the national estimate that is, by definition, an average that misrepresents every population subgroup. In fact, the more variation in skill in the population among population subgroups, the more misleading national averages are.

Projections of literacy and numeracy skill for future periods

The same statistical methods have been used to generate projections of literacy skill distributions for future periods. In this case the relationship between skill and key background characteristics are applied to population projections that include the same variables e.g. age, gender, education. These data provide a means to compare the expected literacy and numeracy skill supply to estimates of the expected economic demand for literacy and numeracy skill implied in the shifting distribution of employment by occupation. These estimates, while useful, can only provide a base case for policy makers as they assume that the relationship between skill and background characteristics included in the model remain stable over the projection period.

A recommendation for adopting synthetic estimation as a first step

It is recommended that less developed countries apply these synthetic estimation methods as a first step towards fielding a full adult skill assessment as the approach reduces the financial and operational burden to acceptable levels and is likely to produce proficiency estimates that are sufficient to support most related policy decisions. Even when the estimates do not support all of their intended uses the data generated can be used to improve the information yield of a “real” assessment in ways that maximize the information yield by optimizing sample design and sample allocation.

Ideally, one would implement synthetic estimation using data collected by a fully-adaptive test such as TOWES-Prime that offers significantly more precise individual estimates, so the imputation models will capture more of the underlying variance in proficiency scores by demographic characteristics.

8. Recommendations for the way forward: A summary

The next cycle of PIAAC is scheduled to be conducted in 2021, a schedule that implies that item development and validation will have to begin in 2018.

It is recommended that countries that have levels of education attainment close to OECD levels consider participation in PIAAC. These countries should argue for the development of larger item pools and the introduction of higher levels of adaptivity in the design so that they get more reliability in that part of the proficiency distribution where most of their population will fall.

For countries with education levels well below those of the OECD member states it is recommended that they field a web-based, fully adaptive skill assessment using the TOWES=Prime assessment or a more adaptive version of the PIAAC-on line assessment.

Proposal of a timeline for the option selected

Two cycles of PIAAC have been conducted that provide data for a significant number of economically advanced countries. The STEP and LAMP studies provide results for a smaller group of low- and middle-income countries.

A. In 2018, produce a first comprehensive set of global estimates

It is recommended that the relationships between assessed skill and key demographic covariates be combined with current data on the same covariates from for each country in the world to produce a projected global distribution of literacy and numeracy skill for 2018. The fact that the reference periods of the two estimates do not line up is not problematic as research on the IALS, ALL and PIAAC data suggests that skill distributions evolve very slowly. The availability of these estimates will guard against the fact that shifting coverage of assessments confounds comparison over time. The publication of these estimates will also create an incentive for countries who have yet to field an assessment to participate in the 2021 collection cycle or one of the alternate strategies recommended in this report.

B. From 2019 onwards, produce an annual comprehensive set of literacy and numeracy estimates

C. In 2019, ensure that a large number of countries participate in PIAAC or one of the alternate assessments

Processes should then be put in place that encourage a large number to either participate in the 2021 cycle of PIAAC or to field one of the less technically, operationally and financially demanding options outlined in this report.

D. From 2019 onwards, produce an annual comprehensive set of literacy and numeracy estimates

The global literacy and numeracy model would then be updated and reported on an annual basis by UIS. Any new assessment data could be integrated as it becomes available. These data would provide national and international users with a regular and current source of evidence, a prerequisite to maintaining policy focus and to adjusting policies and programs. A large-scale rebasing of the model would be undertaken in 2031 when the next PIAAC collection cycle is undertaken.

Several additional recommendations will serve to enhance the utility and salience of new assessment cycles as outlined below.

Build new items to convey a deep understanding of the assessment framework and to improve face validity

Were TOWES-Prime, PIAAC on-line or an new adaptive assessment platform be adopted it is recommended that participating countries receive training in the frameworks and be asked to develop items that are piloted with a view to being added to the international item pool.

Among the benefits of participation in the PIAAC program is related to the fact that officials from participating countries receive training in the theory and approach to measurement used to assess skill and then get to apply these skills to develop test items that are considered for inclusion in the assessment. This process provides national teams with a deep understanding of the proficiency measures and how they should be interpreted.

Experience with the NALS and LSUDA assessments suggests that the approach to measurement employed by PIAAC is somewhat problematic for 15 year olds. It would seem that surprisingly high proportions of these youth have yet to be exposed to the types of authentic tasks upon which PIAAC relies. The associated response dissonance causes them to be placed lower on the scale than would be the case if they were tested with more conventional school-type items. Reports should note this additional source of unreliability.

In keeping with the objective of reducing costs it is recommended that the items be piloted on a real subsample of selected individuals i.e. individuals with known probabilities of selection. Given that the probability that items that are faithful to the frameworks fail to perform once professionally reviewed is very low, and the fact that fully adaptive tests reduce the weight of non-performing/out of pattern test items, in most cases the pilot cases can be retained for use in the final estimates.

Run workshops on how to use the data to set policy and improve practice

The last 30 years have provided an opportunity to understand how the PIAAC proficiency data can be used to inform policy and instructional practice.

For policy, the IALS, ALL, PIAAC and STEP data have been used to document the overall skill levels of populations and to identify population sub-groups whose skill places them at risk. These data have been used to identify the need for additional investment and where additional investment might yield the greatest returns on investment. The same data have also been used to shift the focus of policy strictly from the supply side - i.e. the generation of new skill supply - to include measures that increase skill demand side and that improve the efficiency of the labour market in matching worker skills to the demands of jobs. These latter measures ensure that any skill that gets generated is put to productive use.

On the instructional side, recent research in Canada has established, through the use of large-scale, randomized controlled trials, that adult literacy and numeracy skill upgrading have a significant positive impact on individual and firm outcomes. More directly, increased skill leads to a range of positive benefits to both the employer and the individual provided that the instruction is of high enough quality to generate meaningful levels of skill gain. These returns averaged 25% in the first year, high enough to justify public investment.

Literacy skill gains in these studies have averaged up to 26 points on the 500-point proficiency scale, from as little as 15 hours of construct-focused classroom instruction. This amount of skill gain is equivalent to half a literacy proficiency level and equal to the average skill gain associated with a full additional year of formal education in OECD countries (SRDC, 2013).

Analysis of these same data has also revealed significant failures in the markets for skill generation and utilization that have been traced back to a lack of information and incentives. If not corrected, these failures will attenuate the return on investment in any policy measures designed to upgrade the skill levels of adults. Countries need to be made aware of this research.

Develop measures of occupational skill demand

Although literacy and numeracy are key to both to economic development and social progress the available evidence suggests that the economic dimensions of these skills is central to economic performance. Evaluating the adequacy of current levels of literacy and numeracy for meeting economic objectives depends crucially upon understanding trends in the occupational demand for these skills.

It is recommended that UNESCO and the OECD work with the US government to adapt the O-net skill taxonomy for international use. Application of this taxonomy to national employment by occupation data would provide first order approximations of literacy and numeracy skill demand that can be compared to skill supply estimates with a view to identifying notional skill shortages and surpluses.

Recommended Interim reporting strategy

As noted above all countries should be encouraged to field either PIAAC or a web-based adaptive assessment that is capable of supporting reporting on the PIAAC scales.

Any recommendation about interim reporting strategies must necessarily start with how IALs, ALL, PIAAC and STEP data have been reported.

The first level of reporting has been to publish rank-ordered estimates average scores for the total population and for key population subgroups. The importance of reliable comparison cannot be stressed enough as it provides the foundation for any policy analysis. As noted above, one of the key motivations in fielding an assessment is to understand how ones average scores compare to those of key trading partners. Differences in average literacy scores among countries have been shown to explain fully 55% of long term differences in rates of GDP and labour productivity growth – so a country's average score relative to others matters.

The international comparative reports from IALS, ALL and PIAAC also reported comparative distributions of skill by proficiency level. In IALS and ALL these latter comparisons highlighted the proportions at Level 1 and 2 as analysis of the impact of skill on individual outcomes showed that adults at these proficiency levels shared a disproportionate share of poor outcomes. As noted above, as the PIAAC coverage has expanded to include a significant number of non-OECD countries, the case for a single standard diminished.

It is recommended that international bodies refrain from setting any proficiency level threshold, explicit or implied, in their presentation of proficiency averages and distributions by level. This said, UNESCO should undertake an analysis of the relationship between proficiency level and key democratic, labour market, educational, health and social outcomes. Given the relationships observed in the available national datasets, it is likely that one will see rapid increases in the probability of people experiencing poor outcomes at lower skill levels. It is also likely that the point at which the likelihood of experiencing poor outcomes is the same in all countries or in groups of countries at a similar level of development. If this is the case UNESCO could consider applying the empirical thresholds that imply a level of skill-based disadvantage that offends groups human rights.

It is further recommended that UNESCO encourage member states to undertake national consultations with a view to establishing national skill targets believed to be needed to achieve national economic and social goals, including but not limited to realizing target economic growth rates and to contain skill-based social inequalities in outcomes to acceptable levels.

More specifically, key questions needing national debate leading towards consensus include:

- What is the skill level needed to meet targets for productivity and GDP growth?
- What is the skill level needed to reduce skill-based inequalities to acceptable levels?
- What is the minimum skill level needed to get full returns on public and private investments in tertiary education? ...in health care?

These targets would be published nationally and skill distributions compared to them.

The second level of IALS, ALL, PIAAC and STEP reporting focused on documenting variation in the impact of skill on individual educational, labour market, health and social outcomes. The international comparative reports included analyses that showed how likely people at each skill level were to experience poor outcomes. These analyses included unadjusted estimates of the probability of experiencing poor outcomes and estimates of the probability that were adjusted for the effect of a broad range of other variables known to other influence the probability of experiencing a poor outcome. For example, the unadjusted probability of someone at Level 2 literacy being in fair or poor health might be 5 times higher than someone at Level 5. This probability might fall to 2.5 times once one has removed the effect of age, education, income, gender and mother tongue, occupation and geographic location on the probability. At some level, skill-based inequality will be high enough to consider it to impair people's human rights and welfare. These analyses reveal that much of observed inequality in labour market, educational, health and social outcomes are skill-based. The fact that these differences in life risk have been shown to be causal creates a moral obligation for policy makers to pursue policies that seek to reduce these skill-based inequalities

It is recommended that international bodies report comparative measures of absolute and relative risk associated with each proficiency level. These comparisons will highlight the degree to which inequalities in key outcomes are skill-based and the relative magnitude of the disadvantage.

The third level of IALS, ALL, PIAAC and STEP reporting included analyses of the determinants of skill using cross-sectional estimates of skill reveal the central importance of formal education to skill acquisition. Notwithstanding this general finding, analysis also identifies larger variance in skill at every educational level that is simply too large to attribute to differences in the quality of education. Analysis of the indices of skill use at work and outside work suggest that high level of skill use lead to skill gain over the life course and that low levels of skill use lead to skill loss over the life course.

It is recommended that international bodies publish analyses of the determinants of observed skill levels. It is important, however, that these analyses be accompanied by caveats about the impact that skill gain and loss over the life course are likely to have had on the strengths of the observed relationships.

Synthetic cohort analysis that uses data from multiple assessment cycles, and analysis of true longitudinal data sets confirms significant skill gain and loss over the life course that is related to intensity of skill use and fit between the skill demands of jobs and the skills of workers in those jobs. Adults with skills above the levels demanded by their jobs are at particular risk of losing skill that cost a great deal to acquire in the first place.



It is recommended that countries field a minimum of two assessments to allow for a partial reconstruction of skill gain and loss trajectories by age group and demographic characteristic, information that is crucial to understanding the need to balance skill supply-side policy measures with skill demand-side and market efficiency measures.

Finally, it is recommended that UNESCO develop model-based annual estimates for all countries, including any country not fielding an assessment. The models would use the covariance structure observed in comparator countries to impute score distributions for un-assessed populations. The availability of these estimates will ensure that reported progress over time will reflect more than shifts in the coverage of countries for which data is available



Annex A: References

Neice, D. and Murray, T. S. (1997) The reliability of self-reports of literacy and numeracy skill, UNESCO Institute for Education, Hamburg

Satin, A. and Murray, T.S. (1990) Approaches to literacy and numeracy assessment, UN Statistical Commission, New York

Murray T.S. and Shillington, R. (2013) Understanding Literacy Markets in Alberta: A Market Segmentation Analysis, DataAngel Policy Research, Ottawa

World Bank (2015). Skills Toward Employment and Productivity: Global initiative to generate internationally comparable data on skills of adult populations. Retrieved from http://documents.worldbank.org/curated/en/516741468178736065/122290272_201511338025153/additional/897290NWP0P132085290B00PUBLIC001421.pdf

Gaëlle Pierre, Maria Laura Sanchez Puerta, Alexandria Valerio and Tania Rajadel (2014). STEP Skills Measurement Surveys: Innovative Tools for Assessing Skills. World Bank Group Social Protection and Labour. <http://documents.worldbank.org/curated/en/516741468178736065/pdf/897290NWP0P132085290B00PUBLIC001421.pdf>

Sir Arthur Lewis Institute of Social and Economic Studies (SALISES) (2012). Barbados Country Assessment of Living Conditions 2010 Volume 1: Human Development Challenges in a Global Crisis: Addressing Growth and Social Inclusion. Retrieved from <http://www.caribank.org/uploads/2012/12/Barbados-CALC-Volume-1-MainReport-FINAL-Dec-2012.pdf>

PIAAC Literacy Expert Group (2009), "PIAAC Literacy: A Conceptual Framework", OECD Education Working Papers, No. 34, OECD Publishing, Paris. <http://dx.doi.org/10.1787/220348414075>

UNESCO (2004) "The Plurality of Literacy and its implications for Policies and Programs" (PDF). UNESCO Education Sector Position Paper: 13. 2004.

Annex B Recently applied definitions of literacy

Overview of the conceptual definitions of literacy or functional literacy by agency or programme

Broad definition of literacy	Study (assessments/survey) or national/international documents	Note/Reference
<p>Literacy is using printed and written information to function in society to achieve one's goals and to develop one's knowledge and potential.</p> <p>Both IALS and ALL defines skills along a continuum of proficiency. There is no arbitrary standard distinguishing adults who have or do not have skills. For example, many previous studies have distinguished between adults who are either "literate" or "illiterate". Instead, the ALL study conceptualizes proficiency along a continuum and this is used to denote how well adults use information to function in society and the economy. (Stat Can and OECD, 2011)</p> <p>Four skill assessment domains in ALL</p> <ul style="list-style-type: none"> • Prose literacy – the knowledge and skills needed to understand and use information from texts including editorials, news stories, brochures and instruction manuals. • Document literacy – the knowledge and skills required to locate and use information contained in various formats, including job applications, payroll forms, transportation schedules, maps, tables and charts. • Numeracy – the knowledge and skills required to effectively manage the mathematical demands of diverse situations. • Problem solving – Problem solving involves goal-directed thinking and action in situations for which no routine solution procedure is available. The problem solver has a more or less well defined goal, but does not immediately know how to reach it. The incongruence of goals and admissible operators constitutes a problem. The understanding of the problem situation and its step-by-step transformation based on planning and reasoning, constitute the process of problem solving. (Stat Can and OECD, 2011, p.14) 	<p>The International Adult Literacy Survey (IALS) (conducted in 22 countries between 1994 and 1988)</p> <p>The Adult Literacy and Life skills (ALL) Survey (conducted in 11 countries between 2003 and 2008)</p>	<p>https://nces.ed.gov/surveys/ials/</p> <p>https://nces.ed.gov/surveys/all/index.asp</p> <p>http://www.statcan.gc.ca/pub/89-604-x/89-604-x2011001-eng.pdf</p>
<p>Literacy is understanding, evaluating, using and engaging with written text to participate in the society, to achieve one's goals and to develop one's knowledge and potential</p>	<p>Survey of Adult Skills (PIAAC)</p>	<p>http://www.oecd.org/skills/piaac/</p>

Broad definition of literacy	Study (assessments/survey) or national/international documents	Note/Reference
<p>“ability to identify, understand, interpret, create, communicate and compute, using printed and written materials associated with varying contexts. Literacy involves a continuum of learning in enabling individuals to achieve their goals, to develop their knowledge and potential, and to participate fully in their community and wider society”</p>	<p>UNESCO; UIS Literacy Assessment and Monitoring Programme (LAMP)</p>	<p>http://unesdoc.unesco.org/images/0013/001362/136246e.pdf</p>
<p>Cognitive skills are defined as the “ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, to overcome obstacles by taking thought”</p>	<p>World Bank Skills Towards Employability and Productivity (STEP)</p>	<p>http://microdata.worldbank.org/index.php/catalog/step/about</p>
<p>Literacy refers to the ability to read and write at a level whereby individuals can effectively understand and use written communication in all media (print or electronic), including digital literacy.</p>	<p>European Declaration of the Right to Literacy (European Literacy Policy Network, or ELINET, 2016, p.3)</p>	<p>http://www.elinet.eu/fileadmin/ELINET/Redaktion/user_upload/European_Declaration_of_the_Right_to_Literacy2.pdf</p>
<p>ELINET uses a multi-layered definition of literacy, from baseline literacy to functional and multiple literacy:</p> <p>Baseline literacy means having the knowledge of letters, words and text structures that is needed to read and write at a level that enables self-confidence and motivation for further development.</p> <p>Functional literacy stands for the ability to read and write at a level that enables someone to develop and function in society, at home, at school and at work.</p> <p>Multiple literacy corresponds to the ability to use reading and writing skills in order to produce, understand, interpret and critically evaluate written information. It is a basis for digital participation and making informed choices pertaining to finances, health, etc.</p>	<p>European Literacy Policy Network (ELINET)</p>	
<p>The core skills include language, literacy and numeracy skills of learning, reading, writing, oral communication and numeracy</p>	<p>Australian Core Skills Framework; Foundation Skills Assessment Tool</p>	<p>https://docs.education.gov.au/system/files/doc/other/acsf_document.pdf;</p>

Broad definition of literacy	Study (assessments/survey) or national/international documents	Note/Reference
		https://www.education.gov.au/foundation-skills-assessment-tool
Essential skills are not technical skills but rather skills people use to carry out a wide variety of occupational tasks and daily activities	Canada's Test of Workplace Essential Skills	http://www.eli-net.eu/fileadmin/ELINET/Redaktion/user_upload/Denmark_Short_Report.pdf
<p>Literacy is understood as combination of "skills, tasks, practices, and changes in learners' lives" (OALCF, 2011, p. 3).</p> <p>"Skills are discrete descriptors of literacy and numeracy development, such as decoding, sentence structure, and locating information" (OALCF, 2011, p. 3).</p> <p>Practises: "understanding literacy and numeracy as a social practise involves consideration of what people are doing, feeling, and thinking when engaged with actual print and numeracy activities" (OALCF, 2011, p. 3).</p> <p>"Tasks emphasize more than skills, as they consider purpose, context, and culture to reflect actual use" (OALCF, 2011, p. 3). Change: "people respond to change and make changes in their lives and the lives of others when they participate in a literacy programme" (OALCF, 2011, p. 3).</p>	Canada: Ontario Adult Literacy Curriculum Framework	http://www.tcu.gov.on.ca/eopg/publications/OALCF_Selected_Assessment_Tools_Mar_11.pdf
Literacy as acquiring the skills of reading, writing and arithmetic and the ability to apply them to one's day-to-day life	India National Literacy Mission	http://www.nlm.nic.in/unesco_nlm.htm
The Learning Progressions for Adult Literacy and Numeracy refer to "the main elements or strands of learning adults require in order to listen with understanding, speak to communicate, read with understanding, write to communicate, make sense of number to solve problems, reason statistically and measure and interpret shape and space" (Tertiary Education Commission, 2008a, as cited in Ministry of Education, 2014, p. 3).	Adult Literacy and Numeracy Progressions, Tertiary Education Commission, New Zealand	It is noted that the Assessment Tool is intended for individual testing. It is designed to provide an accurate measure of the skill of each respondent that can be used to inform teaching and learning (Yamamoto, 2013).

Broad definition of literacy	Study (assessments/survey) or national/international documents	Note/Reference
		https://www.educationcounts.govt.nz/_data/assets/pdf_file/0009/114957/Assessing-skills-of-adult-learners-in-2011.pdf
<p>Functional illiteracy exists when the written skills of adults are lower than those which are the minimum and considered a matter of course to cope with day-to-day requirements in society. [...]... If a person cannot read one or several items of information directly contained in a simple text so that the sense is understood and/or is at a similar skills level when writing.</p>	Germany, Level One Study	http://blogs.epb.uni-hamburg.de/leo/files/2011/12/leo-Press-brochure15-12-2011.pdf
<p>Definitions of literacy There are several different definitions that guide policy and practice in France Illettrisme France uses the term "illettrisme" (illiteracy) for persons who have attended school in France for at least 6 years but have not acquired sufficient mastery of reading, written communication, calculation, and other basic competencies to be autonomous in their daily lives. In the policy context of the struggle against illiteracy these individuals need to re-learn and reacquaint themselves with the culture of writing and basic learning. (Jeantheau, 2005, p.77; ANCLI) Analphabétisme "Analphabétisme » is used for people who have never been schooled and « have never learned any written code (most of these people are foreigners or of foreign origin)» They would be entering an initial level of learning. (Jeantheau, 2005, p.77; ANCLI) Français Langue Etrangère (FLE) « People who learn French as a foreign language are generally not considered « analphabètes » and form another specific group within the French classification of people with basic skills needs in the French Language." They are considered second-language learners. (Jeantheau, 2005, p.77; ANCLI)</p>	France' the Information and Everyday Life Survey (or Information et Vie Quotidienne, IVQ)	http://www.centreforliteracy.qc.ca/sites/default/files/IVQ_doc_en.pdf
<p>The Country's national strategy for the encouragement and increase of literacy 2014 - 2020 Strategy refers to the definition of literacy adopted by UNESCO in 2003: "Literacy is the ability to identify, understand, interpret, create, communicate and compute, using printed and written materials associated with varying contexts. Literacy involves a continuum of learning in enabling individuals to</p>	Bulgaria National strategy for the encouragement and increase of literacy 2014 - 2020	http://www.eli-net.eu/fileadmin/ELINET/Redaktion/user_upload/Bulgaria_Adults_Report.pdf

Broad definition of literacy	Study (assessments/survey) or national/international documents	Note/Reference
<p>achieve their goals, to develop their knowledge and potential, and to participate fully in their community and wider society.”</p> <p>The Strategy presents an analysis of the factors and trends for the development of literacy, provides a definition for the various types of literacy and clarifies the scope and content of the concept of "literacy".</p> <p>Depending on the functions, objectives and age groups to which they relate, the following types of literacy are distinguished:</p> <ul style="list-style-type: none"> • basic literacy – skill of reading and understanding a text, writing and proper use of the language in a particular context (related to the competences which students are expected to possess at the end of grade IV); • functional literacy – a skill of discovering, selecting, extracting and synthesising information from various sources and its use to achieve a particular goal both in learning of all subjects of study and in various real-life situations (it is related to the competences which the students are expected to possess at the end of their basic education); • multifunctional literacy – ability to create, understand, interpret and critically assess written information (it is related to competences expected to be developed and demonstrated within training for acquisition of secondary education and of higher education degrees). 		
<p>The National Adult Literacy Agency (NALA) defines literacy as listening and speaking, reading, writing, numeracy and using everyday technology to communicate and handle information. Literacy includes more than the technical skills of communication: it also has personal, social and economic dimensions. Literacy increases the opportunity for individuals and communities to reflect on their situation, explore new possibilities and initiate change (Government of Ireland, 2012).</p>	<p>Ireland National Adult Literacy Agency (ELINET, 2016, p.13)</p>	<p>http://www.elinet.eu/fileadmin/ELINET/Redaktion/user_upload/Ireland_Adults_Report.pdf</p>