Organisation for Economic Co-operation and Development

EDU/WKP(2020)28

Unclassified **English text only**

11 December 2020

DIRECTORATE FOR EDUCATION AND SKILLS

Cancels & replaces the same document of 10 December 2020

IMPROVING THE COLLECTION OF INFORMATION ON LITERACY PROFICIENCY IN HOUSEHOLD SURVEYS

OECD Education Working Paper No. 240

By William Thorn, OECD

This working paper has been authorised by Andreas Schleicher, Director of the Directorate for Education and Skills, OECD.

William Thorn, William.Thorn@oecd.org

JT03469704

OECD EDUCATION WORKING PAPERS SERIES

Comment on the series is welcome, and should be sent to edu.contact@oecd.org.

This working paper has been authorised by Andreas Schleicher, Director of the Directorate for Education and Skills, OECD.

---

www.oecd.org/edu/workingpapers

Abstract

In the vast majority of the world's countries, information on the literacy proficiency of the adult population is collected through census collections, labour force surveys or through omnibus household surveys. These commonly use simple measures: respondents' reports of their own or other household members' capacity to read and write or the capacity of the respondent to accurately read aloud a short sentence.

While there is a justified interest in the use of assessments to collect information regarding literacy proficiency, household surveys using simple measures will continue to be a primary source of data on literacy in many countries for some time. Improvement of the quality of simple measures should, therefore, be a priority. Three main avenues for improvement are identified: greater clarity regarding the concepts being measured, the development of improved simple direct assessments of literacy proficiency and encouragement for the use of a common set of instruments and questions.

Résumé

Dans la grande majorité des pays, les données sur le niveau de compétence en littératie des adultes sont collectées grâce à des recensements, des enquêtes sur la population active ou des enquêtes omnibus auprès des ménages. Ces enquêtes utilisent généralement des mesures simples. Elles sont basées sur les déclarations des répondants à propos de leur capacité, ou celle d'un autre membre du ménage, à lire et à écrire, ou bien sur la capacité du répondant à lire correctement à voix haute une phrase courte.

Aujourd'hui, le recours aux évaluations pour collecter des informations sur les compétences en littératie suscite, certes, un intérêt légitime mais les enquêtes auprès des ménages qui utilisent des mesures simples continueront encore longtemps à constituer la principale source de données sur la littératie dans de nombreux pays. L'amélioration de la qualité de ces mesures simples doit donc figurer au rang des priorités. Trois pistes principales ont été identifiées à cet effet : l'amélioration de la clarté des concepts mesurés, l'amélioration des évaluations directes simples des compétences en littératie, et l'utilisation plus fréquente d'un ensemble commun d'instruments d'évaluation et de questions.

Table of Contents

Tables

Figures

Boxes

Abbreviations and acronyms

ALL	Adult Literacy and Life Skills Survey
BCS	British Cohort Study
DHS	Demographic and Health Surveys
EFA	Education for All
EPDC	Education Policy and Data Center
ETS	Educational Testing Services
IALS	International Adult Literacy Survey
IHSN	International Household Survey Network
ISCED	International Standard Classification of Education
LAMP	Literacy Assessment and Monitoring Programme (UNESCO)
LSMS	Living standards measurement surveys (World Bank)
MICS	Multi-indicator cluster surveys (UNICEF)
OECD	Organisation for Economic Co-operation and Development
PASEC	Programme d'analyse des systèmes éducatifs de la Confemen (Conférence des ministres de l'Éducation des États et gouvernements de la Francophonie – Programme for the analysis of educational systems of the Conference of Education Ministers of States and governments of the Francophonie.
PIAAC	Programme for the International Assessment of Adult Competencies (OECD)
SDGs	Sustainable development goals
STEP	Skills Towards Employability and Productivity Measurement Program (World Bank)
UIS	UNESCO Institute for Statistics
UNESCO	United Nations Educational, Scientific and Cultural Organization

Acknowledgements

This paper was written by William Thorn, Senior Analyst in the OECD's Directorate for Education and Skills. Vanessa Denis and François Keslair provided research and statistical assistance. Jennifer Cannon and Sabrina Leonarduzzi prepared the paper for publication.

The preparation of this paper benefited from review and comments by Bryan Maddox, Anke Grotlüschen, Irwin Kirsch, Greg Brooks and Felix Singleton Thorn as well as colleagues in the OECD Directorate for Education and Skills.

Improving the collection of information on literacy proficiency in household surveys

1. Introduction

In the vast majority of the world's countries, information on the literacy proficiency of the adult population is collected through census collections, labour force surveys or through omnibus household surveys such as UNICEF's multi-indicator cluster surveys (MICS), the demographic and health surveys (DHS) programme financed by US AID and the World Bank's living standards measurement surveys (LSMS).[1] These types of studies commonly use simple measures that do not take much time to administer: respondents' reports of their own or other household members' capacity to read, write and calculate, or very short tests such as the assessment of the capacity of the respondent to accurately read aloud a short sentence.

At the international level, there is a clear intent to support the development and wider use of assessments as the vehicle for the collection of data on literacy, in large part driven by the approach chosen for the definition and measurement of the sustainable development goals (SDGs) relating to education and literacy. While the number of countries carrying out large-scale assessments is likely to increase over the next few years, it can nevertheless be expected that for reasons of cost, complexity and capacity, many countries, especially low- and middle-income countries will not be in a position to implement such assessment studies in the short to medium-term.

In this paper, a distinction is made between two types of direct measures (tests) of literacy proficiency*: literacy assessments* and *simple direct measures* of literacy. The term *assessment* is used to refer to relatively lengthy and complex tests designed to accurately describe the proficiency of test-takers on a continuous scale. The term "*simple direct measures*" is used to refer to very short tests designed to locate individuals in terms of a categorical classification with a limited number of categories (usually 2-4).

The argument developed below is: (1) that household surveys using simple measures will continue to be a primary source of data on literacy in many countries for many years to come; and (2) that there is considerable scope to improve the quality of data on literacy (and numeracy) collected in household surveys through these types of measures. Three main avenues for improvement are identified: greater clarity regarding the concepts being measured, the development of improved simple direct measures of literacy proficiency and encouragement for the use a common set of instruments

[1] Information on the sources of literacy data is available in a file downloadable from the UNESCO Institute of Statistics (UNESCO Institute for Statistics (UIS), n.d.[3]).

and questions for the collection of information on literacy (and numeracy) across studies.

Four broad categories of information about literacy are collected in household surveys that cover this topic: information on (1) literacy proficiency; (2) literacy practices; (3) participation in literacy programmes; and (4) contextual or environmental factors related to the development and maintenance of literacy proficiency such as access to libraries or the presence of books in the family home. This paper concentrates exclusively on the measurement of proficiency as this is the focus of the SDGs and is the concept most commonly measured. This should not be seen as a judgement regarding the relative importance of information about proficiency and information about other dimensions of literacy. This latter information is also extremely important for an understanding of the context in which literacy proficiency is developed and maintained in different countries. Should a set of international guidelines for the collection of information on literacy in household studies be developed as is recommended in this paper, coverage of these other dimensions of literacy would be essential.

The structure of this paper is the following. Section 2 provides the context for the paper, describing the Sustainable Development Goals (SDGs) and experience with international literacy assessments. Section 3 describes the information that is collected regarding proficiency in censuses and omnibus household surveys. This information is reviewed in Section 4, concentrating on issues of validity and comparability and is followed by a concluding section (Section 5) which develops some recommendations for improving the collection of information on literacy.

2. Background

The Sustainable Development Goals and the measurement of literacy

In September 2015, the General Assembly of the United Nations agreed to an Agenda for Sustainable Development which established 17 goals and 169 targets designed to mobilise global efforts to end all forms of poverty, fight inequalities and tackle climate change (United Nations (UN), 2015[1]). These replaced the Millennium Development Goals and the Education for All goals. The goals are operationalised in the form of targets with accompanying statistical indicators against which progress towards the goals is to be measured.

Goal 4 of the SDGs relates to education and training and is to: "Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all." Among the targets defined for Goal 4 is Target 4.6: "By 2030, ensure that all youth and a substantial proportion of adults, both men and women, achieve literacy and numeracy." The indicator (4.6.1) defined to measure progress towards this target is the "[p]roportion of the population in a given age group achieving at least a fixed level of

proficiency in functional (a) literacy and (b) numeracy skills, by sex."[2] This is further specified in the following way (UN Statistics Division (UNstats), n.d.[2]):

> *The proportion of youth (aged 15-24 years) and of adults (aged 15 years and above) who have achieved or exceeded a given level of proficiency in (a) literacy and (b) numeracy. The minimum proficiency level will be measured relative to new common literacy and numeracy scales currently in development.*

Indicator 4.6.1 is explicitly conceived as "a *direct measure* [emphasis added] of the skill levels of youth and adults. The fixed level of proficiency is the benchmark of basic knowledge in a domain (literacy or numeracy) measured through *learning assessments* [emphasis added]" (UN Statistics Division (UNstats), n.d.[2]).

Defining the indicator for Target 4.6 in this way assumes that comparable data from literacy and numeracy assessments will be available for a large share of the countries in the world by 2030. These data would need to come from either the use of common assessment instruments or the use of assessment instruments that could be linked or equated in some way. In order to examine whether this assumption is reasonable, the sources of current data on literacy are reviewed in the next section as plans for the implementation and further development of literacy assessments over the next decade.

International measures of literacy: The current situation

Information on the literacy proficiency of the adult population comes predominantly from three sources:

- Population censuses
- Household surveys
- Dedicated literacy assessments.

Household surveys

In the majority of countries, the source of data on literacy is generally population census or household surveys such as labour force surveys or surveys conducted as part of the MICS, DHS and LSMS programmes (UNESCO Institute for Statistics (UIS), n.d.[3]). The information on literacy proficiency from these sources consists of responses to questions concerning the respondent's and/or other household members' capacity to read and write or responses to a very simple reading "test". In addition, the collection of this information is also often restricted to a subset of the target population. For example, both the MICS and DHS questionnaires only collect direct information relating to literacy proficiency (i.e. whether respondents can read) from respondents who have no schooling or have attended primary schooling only. Respondents who have attended

[2] The term 'functional' literacy (and numeracy) is used in the description of the indicator, but nowhere else in the documentation for Goal 4.6.

secondary level education or higher are assumed to be literate. Very few high-income countries collect information on literacy from these types of sources. Given universal or near universal literacy in these countries, there is little value in collecting information on whether adults can read and write.

Literacy assessments

Dedicated literacy assessments are undertaken in a minority of countries. Five international assessments of adult literacy have been implemented since the mid-1990s: the International Adult Literacy Survey (IALS), the Adult Literacy and Life Skills Survey (ALL), the Programme for the International Assessment of Adult Competencies (PIAAC), UNESCO's Literacy Assessment and Monitoring Programme (LAMP) (UNESCO Institute for Statistics (UIS), 2017[4])) and the World Bank's STEP measurement study (Gaëlle et al., 2014[5]). Of these projects, PIAAC is the only project currently in operation. A list of the countries participating in these assessments is provided in Table 1.

It should be noted that while the target population for IALS, ALL, PIAAC and LAMP was generally the national population of working age (within the age range of 15/16-64/65 years), the target population of STEP is generally the population in major urban centres, not the national population. In addition to these international assessments, national assessments of adult literacy exist in several countries. Examples include Bangladesh, Botswana, England (United Kingdom), France, Germany, Kenya, Scotland (United Kingdom) and the United States.

Table 1. Participation in international literacy assessments

IALS 1994-98	ALL 2003-07	PIAAC (Cycle 1) 2008-19	PIAAC (Cycle 2) 2018-24	LAMP 2003-2011	STEP[2] 2012-2017
Australia	Australia	Australia	Australia	Jordan	Armenia
Canada	Bermuda	Austria	Austria	Mongolia	Azerbaijan
Chile	Canada	Canada	Canada	Palestine	Bolivia
Czech Republic	Italy	Chile	Chile	Paraguay	Colombia
Denmark	Hungary	Cyprus[1]	Croatia		Georgia
England (UK)	Netherlands	Czech Republic	Czech Republic		Ghana
Finland	New Zealand	Denmark	Denmark		Kenya
Flanders	Norway	Ecuador	England (UK)		Kosovo
Germany	Nuevo Leon (Mexico)	England (UK)	Estonia		Lao People's Democratic Republic
Hungary	Switzerland	Estonia	Finland		North Macedonia
Ireland	United States	Finland	Flanders (Belgium)		Philippines
Italy		Flanders (Belgium)	France		Serbia
Netherlands		France	Germany		Sri Lanka
New Zealand		Germany	Hungary		Ukraine
Northern Ireland (UK)		Greece	Ireland		Viet Nam
Norway		Hungary	Israel		Yunnan (People's Republic of China)

IALS 1994-98	ALL 2003-07	PIAAC (Cycle 1) 2008-19	PIAAC (Cycle 2) 2018-24	LAMP 2003-2011	STEP[2] 2012-2017
Poland		Ireland	Italy		
Slovenia		Israel	Japan		
Sweden		Italy	Korea		
Switzerland		Japan	Latvia		
United States		Kazakhstan	Lithuania		
		Korea	Netherlands		
		Lithuania	New Zealand		
		Mexico	Norway		
		Netherlands	Poland		
		New Zealand	Portugal		
		Northern Ireland (UK)	Russian Federation		
		Norway	Singapore		
		Peru	Slovak Republic		
		Poland	Spain		
		Russian Federation	Sweden		
		Singapore	Switzerland		
		Slovak Republic	United States		
		Slovenia			
		Spain			
		Sweden			
		Turkey			
		United States			

Notes:
1. Note by Turkey:
The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".
Note by all the European Union Member States of the OECD and the European Union:
The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.
2. The target population for STEP is 15-64 year-olds in major urban centres in the country/region concerned.

Looking forward, development of the second cycle of the OECD's PIAAC began in 2018 with data collection being planned for 2022/23 and the release of results in 2024. Thirty-three countries (all high-income) are participating. No further data collection is planned for LAMP[3] at this stage and plans for STEP are uncertain. In terms of national assessments, Germany administered its Level-One (LEO) study in 2018 (Grotlüschen et al., 2020[6]).

Participation of low and middle-income countries in large-scale literacy assessments faces three main challenges: those of costs, technical capacity

[3] UIS (2018, p. 136[72]) and UIS (2019, p. 34[71]) mention the development of a version of LAMP called mini-LAMP. However, little if any concrete information is available in the public domain about this project.

and making effective use of the data. The costs of participating in international studies such as PIAAC are relatively high as are the technical and organisational demands. Without significant donor support and on-going technical assistance over the six-year implementation period, participation in PIAAC would be difficult for most low- and middle-income countries. Making use of the data collected is also a significant challenge both from the point of view of the existence of local analytical expertise and of the capacity of governments and public authorities to invest in analysis and to integrate results in policy and decision making. If the data are not going to be used or only going to be used for purposes such as reporting on the SDGs, the costs of undertaking such studies will far outweigh the benefits.

There are few ways of reducing the cost and technical barriers to the participation of low and middle-income countries in international adult assessments such as PIAAC without, at the same time, reducing data quality. A significant component of project costs relates to data collection. These are relatively high for two main reasons. First, a household methodology is used involving the presence of the interviewer in the household for between two to four hours. Second, sample sizes are relatively large for both the field test and the main study and a nationally representative sample is required. Costs relating to processes such as international project management, quality assurance and control, data processing, data scaling and analysis are also significant. For example, the international costs for countries participating in the 2nd and 3rd rounds of PIAAC (which involved no development costs) were around EUR 600 000 per country in total. The scope for significant cost savings in implementation is relatively limited without making some compromises concerning data quality and population coverage. In part for reasons of cost, STEP, for example, (1) does not conduct a field test, and (2) in many countries, restricts its target population to the population in major urban centres. However, this has potentially negative consequences in terms of lower data quality due to failure to identify problems with translation, instrument construction and field work procedures at the field test stage and the absence of data at the national level.

The technical barriers to participation include limited in-country expertise relevant to the implementation of large-scale assessments in household settings and the complexity of such studies. In addition to that of STEP, the experience of implementing PIAAC in countries such as Ecuador, Indonesia,[4] Kazakhstan and Peru show that such assessments can be successfully undertaken [see OECD (2019[7]) for results from Ecuador, Kazakhstan and Peru]. However, it is often difficult to maintain stable high quality national teams over the life of the project and that considerable additional support for such countries is required. The experience of LAMP was similar. “The countries taking part lacked expertise in large-scale

[4] Indonesia participated in the Round 2 of PIAAC, collecting data in the Jakarta municipal area only. Results were published in (OECD, 2016[67]). The data were subsequently withdrawn by the Indonesian authorities from the PIAAC public use data files.

assessment and considerable capacity building was required. Sampling and test administration presented considerable logistical challenges" (UNESCO Institute for Statistics (UIS), 2017, p. 148[4]). The provision of significant on-going technical support and assistance represents one option for overcoming these operational challenges. However, this adds to implementation costs. Another option is to reduce reliance on national expertise to the minimum possible. For example, the World Bank managed the data collection process in STEP by contracting directly with a data collection agency in each country concerned.

The idea that there may be ways of gaining information on literacy through approaches to assessment that are "smaller, cheaper and quicker" than those used in large-scale assessments has been promoted by Wagner (2005[8]) and (2011[9]) and Gal (2016[10]). As noted above, collecting data in ways that are "cheaper" necessarily involves trade-offs in terms of the amount of information collected and the quality of data. The proposals of Wagner and Gal are a case in point in that they all involve smaller sample sizes, shorter tests, less sophisticated designs and analysis, less emphasis on quality control, etc. than is the case in studies such as PIAAC. The question is at what point the loss of data quality is such that the data produced are no longer fit for purpose. This point is probably reached far more rapidly than most people imagine.[5] In particular, commentators often ignore the fact that the data collection and field operations phases are among the most important sources of error in survey research. Field operations is also one of the main drivers of costs and, therefore, an area highly likely to be cut back in order to lower budgets.[6]

In summary, it is possible and indeed likely, that the number of countries participating in international literacy assessments or implementing national assessments will increase over the next decade. However, in any realistic scenario, most of the countries in the world will not be in a position to undertake such assessments in the short to medium term. Thus, omnibus household surveys and census collections will continue to be an important, if not the main, source of information on literacy proficiency in many countries in the world, especially low- and middle-income countries, for many years to come. While the promotion of the use of assessments and the mobilisation of funds to support countries that wish to implement literacy assessments should be a key component of a strategy to improve the quality of data on literacy globally, actions to maximise the quality and utility of data on literacy from household surveys and population censuses

[5] Gal acknowledges the existence of these compromises, even if he minimises their significance. The approach he proposes "will of course create scores that are somewhat less reliable and valid compared to the proficiency estimates created by PIAAC or STEP, which use a longer test and more sophisticated statistical procedures" (Gal, 2016, p. 25[10]). "*Somewhat* less reliable and valid" is something of an understatement. "*Considerably* less reliable and valid" would be more accurate.

[6] Regarding the prospects for other methods of sampling (such as non-probability designs) and data collection (e.g. use of mobile devices and mixed- or multi-mode strategies) to reduce costs without compromising data quality, see Link (2018[66]).

should not be ignored. This is an important issue in itself as well as in the context of the efforts to monitor progress towards the SDGs. The following sections of this paper review the measurement of literacy proficiency in household studies.

3. Indirect and simple direct measures of literacy in household surveys

Introduction

This section provides an overview of indirect measures and simple direct measures of literacy implemented in household surveys. The objective is to exemplify the types of measure used. The information about these studies has been drawn largely from the documentation available in the World Bank's Micro-data catalogue and the International Household Survey Network (IHSN) Survey Catalogue as well as the MICS and DHS programme websites.

As noted above, a distinction is made between two types of direct measures (tests) of literacy proficiency*: literacy assessments* (relatively lengthy and complex tests designed to accurately describe the proficiency of test-takers on a continuous scale) and *simple direct measures* of literacy and numeracy (very short tests, sometimes consisting of only one item, designed to locate individuals in terms of a categorical classification).

Indirect measures (respondent reports)

Respondent reports include both self-reports (i.e. the respondent provides information about his/her own literacy proficiency) and third-party reports or reports of others' literacy proficiency (usually that of other household members or other residents of the sampled dwelling unit). Two types of question are found in the studies reviewed. The first is focussed on whether the respondent or household member can read or not. The second focusses on how well the respondent (or household member) can read. The exact wording of questions and the response categories used varies between studies as does the population to which the questions are asked. Typically, two (yes/no) or three (no, yes with difficulty, yes without difficulty) response categories are used for questions in the first group. For questions focussed on how well a person can read, four or five response categories tend to be used (e.g. "cannot read", "poor", "fair", "good", "very good"). There is overlap between the two approaches. Depending on the response categories, questions belonging to the first group may allow some differentiation of the proficiency of those respondents who can read and the questions belonging to the second group may allow differentiation between readers and non-readers. Examples of the questions regarding literacy used in a range of household surveys can be found in Annex B.

Simple direct measures

There are several examples of surveys and survey programmes that use short (often very short) tests to collect information on reading proficiency. The main examples are the MICS and DHS programmes and the World

Bank's STEP measurement study. The Ghana socio-economic panel survey of 2009 (Institute of Statistical Social and Economic Research - University of Ghana, 2009[11]) (part of the World Bank's LSMS programme) also used two short literacy and numeracy assessments. The direct measures used in these studies are described below.

MICS and DHS

The MICS and DHS studies measure the respondent's level of literacy through a simple one-item test. The respondent is asked to read a simple sentence from a show card with the interviewer recording whether the respondent can read the complete sentence, part of the sentence or not read the sentence at all. This item is administered only to those respondents with a highest level of education *at primary level or less*. Respondents with higher levels of educational attainment are *assumed* to be able to read. The exact question format and the response categories can be found in Box 1.

Box 1. Single Sentence Reading Test, Multi-Indicator Cluster Surveys (MICS) and Demographic and Health Surveys (DHS)

Now I would like you to read this sentence to me.

SHOW CARD TO RESPONDENT.

IF RESPONDENT CANNOT READ WHOLE SENTENCE, PROBE: Can you read any part of the sentence to me?

The response categories are:

1. CANNOT READ AT ALL
2. ABLE TO READ ONLY PART OF THE SENTENCE
3. ABLE TO READ WHOLE SENTENCE
4. NO CARD WITH REQUIRED LANGUAGE
5. BLIND/VISUALLY IMPAIRED

In the case that no card with the required language is available, the language in question is recorded by the interviewer.

The sentences used in the MICS are the following (MICS, undated):

- The child is reading a book.
- The rains came late this year.
- Parents must care for their children.
- Farming is hard work.

.*Source*: MICS (n.d.[12]), *MICS6, Instructions for Interviewers*, UNICEF, http://mics.unicef.org/files?job=W1siZiIsIjIwMTcvMDcvMTkvMjAvNDcvMTMvNDY4L01JQ1M2X0luc3RydWN0aW9uc19mb3JfSW50ZXJ2aWV3ZXJzXzIwMTcwNzE5LmRvY3giXV0&sha=667ed1ad05dfc60d (accessed on 6 August 2018).

STEP

The World Bank's STEP measurement study included a literacy assessment that was administered either in the form of a "partial" module or a "full" module (Educational Testing Services (ETS), 2014, p. 7[13]). The partial module consists of an assessment of reading components and eight reading comprehension items known as the literacy core. The full module contains in addition to the assessment of reading component and the core, a full reading assessment in which test-takers take one of four test booklets containing 18 items.

The assessment of reading components lasts around ten minutes on average and is designed to assess whether respondents possess the basic knowledge and skills required for effective reading comprehension. It covers the domains of: (1) print vocabulary; (2) understanding of the semantic logic of sentences (sentence processing); and (3) passage fluency.[7] The core test contains eight items of low difficulty (six are located at or below level 1 and two at level 1 on the PIAAC literacy scale) and takes around seven minutes to complete. In the full module, the core serves to determine whether a test-taker has the ability to continue to the full assessment on not. Those test-takers who "pass" the core (i.e. get at least three of the eight items correct) go on to take the next stage in the assessment while those who "fail" quit the assessment. In the partial module, the core serves to give some basic information about the literacy proficiency of respondents.

An example of the type of an item included in the STEP core is presented in Box 2.

Box 2. STEP core item (modified example)

Question: Here is an advertisement for employment. Circle the number of drivers the delivery company wants to hire.

DELIVERY COMPANY

Wants to hire 5 drivers for immediate start. Must have a valid driver's licence. Part-time work available.
For more information, call:
018 546 794

Source: Modified version of an item from the Skills Towards Employability and Productivity (STEP) literacy core.

[7] The vocabulary items involve the identification of the word that corresponds to a picture. The sentence processing items involve indicating whether or not a sentence (e.g. "The sky is green.") makes sense. Passage fluency items involve reading a passage of text and responding to cloze type items in which the respondent has to select the word that is appropriate in the context. from the two options available (OECD, 2016, p. 23[67]).

Ghana socio-economic panel, 2009

The Ghana socio-economic panel survey (Institute of Statistical Social and Economic Research - University of Ghana, 2009[11]) contained a short literacy and assessment administered to 9-26 year-olds consisting of a version of the single sentence reading test and a short reading comprehension test.

The literacy test involved respondents reading a short text on a show card and answering eight multiple choice questions about the text (Box 3).

Box 3. Short reading comprehension test, Ghana socio-economic panel, 2009

John is a small boy. He lives in a village with his brothers and sisters. He goes to school every week. In his school there are five teachers. John is learning to read at school. He likes to read very much. His father is a teacher, and his parents want him to become a school teacher too.

Who is John?

(a) An old man; (b) A small boy; (c) A school teacher; (d) A school

Where does John live?

(a) In a village; (b) In a city; (c) In a school; (d) In a forest

What does John do every week?

(a) Works with his father; (b) Plays with his friends; (c) Helps his brothers and sisters; (d) Goes to school

How many teachers are there at John's school?

(a) One; (b) Three; (c) Five; (d) Six

What is John doing at school?

(a) Helping the teacher; (b) Talking with his friends; (c) Learning to read; (d) Teaching the class

Who is a school teacher?

(a) John; (b) John's father; (c) John's brother; (d) John's mother

What do John's parents want him to do?

(a) Go to school; (b) Learn to read; (c) Obey his teachers; (d) Become a teacher

The best title for this story is

(a) John Learns to Read; (b) Why Reading is Important; (c) John's Village; (d) Schools in Ghana

Source: Institute of Statistical, Social and Economic Research - University of Ghana (2009[11]) *Ghana - Socio-economic Panel Survey*, http://microdata.worldbank.org/index.php/catalog/2534 (accessed on 6 August 2018).

4. A review of indirect and simple direct measures of literacy proficiency in household surveys

This section, reviews the indirect and simple direct measures of literacy proficiency used in household surveys. The focus is on two questions: the validity of the measures and comparability across studies. In terms of the assessment of validity,[8] the emphasis is placed on the extent to which the measures used provide information about the construct of interest and also on convergent validity. In terms of comparability, the question of interest is the extent to which the measures covering the same constructs or concepts used in different studies are related.

Defining literacy

In examining the validity of the measures of proficiency used in the household studies reviewed, the focus is on the extent to which they can be seen as measures of literacy understood as *the possession of a basic level of reading comprehension or the capacity to construct meaning from (short and simple) written texts and a basic level of ability to produce (short and simple) written texts with communicative intent*. This reflects the discussion of "literacy" in the UN *Principles and Recommendations for Population and Housing Censuses*.

> *Literacy has historically been defined as the ability both to read and to write, distinguishing between "literate" and "illiterate" people. A literate person is one who can both read and write, with understanding, a short, simple statement on his or her everyday life. An illiterate person is one who cannot, with understanding, both read and write such a statement. Hence, a person capable of reading and writing only figures and his or her own name should be considered illiterate, as should a person who can read but not write as well as one who can read and write only a ritual phrase that has been memorized. However, a more modern understanding referring to literacy as a continuum of skills, levels, domains of application and functionality is now widely accepted (United Nations (UN), 2015, p. 236[14]).*

In terms of the reading dimension of literacy, the emphasis placed on the comprehension and understanding of text also reflects the widespread view that "[t]he ultimate goal of learning to read is comprehension" or "the process of simultaneously extracting and constructing meaning through interaction and involvement with written language" (RTI International, 2015, p. 19[15]).[9] This is expressed in the definitions of (reading) literacy underpinning many international comparative assessments (Box 4).

[8] In the words of AERA (2014, p. 11[68]): "Validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests".

[9] See also PASEC (2016[70]): "La compréhension de l'écrit est la finalité de la lecture et implique d'avoir préalablement automatisé les activités de décodage de mots isolés pour accéder progressivement au sens de la phrase puis du

Box 4. Definitions of reading literacy

PIRLS 2016

"Reading literacy is the ability to understand and use those written language forms required by society and/or valued by the individual. Readers can construct meaning from texts in a variety of forms" (Mullis and Martin, 2015[16]).

PIAAC

Literacy is "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" (OECD, 2016, p. 19[17]).

PISA 2009

Reading Literacy is the capacity to "understand, use, reflect on and engage with written texts in order to achieve one's goals, to develop one's knowledge and potential and to participate in society" (OECD, 2010, p. 14[18]).

Literacy: A dichotomy or a continuum?

The UN definition of literacy cited above gives the impression that literacy can be conceived in one of two diametrically opposed ways: as a dichotomy (a person is either literate or illiterate) or as a continuum (people can have more or less literacy). This rather simplifies the issue. It is important to distinguish between: (1) dichotomous *concepts* of literacy; (2) dichotomous *measures* of literacy; and (3) the dichotomous *reporting* of literacy performance.

A dichotomous concept of literacy is one in which being literate is defined purely in opposition to being illiterate and in which any differences in the literacy proficiency of individuals who are literate are seen as unimportant or ignored. Such a concept of literacy has probably rarely if ever been held.

Dichotomous measures are measures with two response categories (e.g. can read or cannot read). These are commonly used (see Annex B). The fact that a dichotomous measure is used in a study should not be interpreted as implying an underlying conception of literacy as a dichotomy. In many, if not most, cases the reasons for the use of binary measures are likely to be pragmatic rather than theoretical – e.g. limited questionnaire space, cost, the likely uses of the information, ease of administration, etc. – or because such a measure is seen as giving sufficient information for the purposes of the study.

Dichotomous reporting of literacy is the reporting of information in a binary format. Multi-category and continuous measures of literacy (similar

texte." "The comprehension of written texts is the ultimate goal of reading and depends on the prior development of the automatic capacity to decode individual words on which is built the progression from the understanding of the meaning of sentences to that of texts". [Author's translation.]

to any other continuous variable such as income or age) can be (and often are) reported in terms of binary classifications: i.e. the proportion of the population above or below a particular threshold value. The indicators proposed for reporting progress towards the literacy goal in the SDGs represent a case in point. As noted above, indicator 4.6.1 presents the proficiency of the population in terms of a *dichotomy*: the proportion of youth and adults who have achieved or exceeded a given level of proficiency in (a) literacy and (b) numeracy. Reporting literacy in this way represents a choice regarding the presentation of the data rather than a consequence of accepting an underlying dichotomous concept of literacy. This may or may not be based on the view that the threshold distinguishing the two reporting categories represents a particularly meaningful point of the scale (i.e. a minimum desired level of performance In most large-scale assessment of literacy, results are reported in terms of a limited number of "levels" to facilitate comparisons and interpretation. Use of a two level rather than multi-level reporting framework represents the "degree zero" of such an approach.

The decision to report using a binary classification is often based on a view that the reporting threshold used has a particular importance – e.g. that it separates members of the population who have an "adequate" or "desirable" level of literacy from those who do not. The view taken in the reporting of the results from IALS that level 3 in prose, document and quantitative literacy represented the suitable minimum level of literacy for "coping with the demands of modern life and work" (OECD/Statistics Canada, 2000, p. 13[19]) represents an example of this as does the SDG approach in which the "fixed level of proficiency is the benchmark of basic knowledge in a domain" (UN Statistics Division (UNstats), n.d.[2]). Maddox and Esposito (2011[20]) offer a good discussion of the problems with and the consequences of reporting literacy in terms of dichotomies.

While there is little doubt that more nuanced measures and reporting categories provide a more accurate representation of the complex reality that is literacy, there is certainly no reason to see the use of binary measures or the binary reporting of literacy proficiency as being incompatible with a conception of literacy as a complex competence or skill which can be mastered more or less well. There may be good reasons to criticise binary measures and binary reporting of results for simplifying reality. However, for some purposes and in some contexts, their use is entirely appropriate.

What is measured by respondent reports?

There is little if any documentation regarding the concepts of literacy measured by the respondent reports of literacy skills in the studies reviewed. This is true even in the case of IALS and STEP where the construct of literacy that is the object of the direct assessment components of these studies is comprehensively described. Questions focussed on determining whether or not the person of interest (respondent or household member) can or cannot read or write can be interpreted as implying a concept of literacy as reading with understanding or writing with communicative intent. However, there is considerable variation in:

- The extent to which the "understanding" or "comprehension" threshold is explicitly specified.
- When the "understanding" or "comprehension" threshold is specified, the way in which this is done.

As can be seen from Annex B, in some surveys, respondents are asked whether they are literate or can read or write without further specification. In others, reading and writing are further specified in terms of the reading (or writing) of certain text and document types, reading and writing in certain languages and reading and writing with understanding or for certain purposes. In summary, even if the underlying concept is the same or similar, the operationalisation of this varies considerably.

In the case of questions in which the focus is on the evaluation of the reading or writing skills of the respondent or a household member (i.e. how well the respondent can read or write), two orientations can be seen. The first represents a broad (non-contextualised) self-evaluation of reading and writing skills. The second represents a self-evaluation of the adequacy of reading and writing skills in a defined context or relative to some defined purpose. Even in the case of IALS, where the construct of literacy that is the object of the direct assessment is described in detail.

What is measured by the single sentence reading test?

There is little information regarding the simple sentence reading test used in the DHS and MICS and the concept of literacy it is intended to measure provided in the published documentation for these studies. The test appears to be intended to measure literacy understood as reading with a minimum level of comprehension and understanding. The MICS interviewer manual presents this item as being designed to "ascertain whether women are literate or not" (MICS, n.d., p. 70[12]). The DHS manual also presents the test as a measure of "literacy" (ICF, 2017, p. 61[21]). In the case of both DHS and MICS, the simple sentence reading test was introduced the late 1990s; in MICS3 and in Phase 4 of DHS (1997-2003). It replaced a question regarding the capacity of household members/respondents to read a letter or a newspaper.[10] The available documentation suggests that the two approaches were intended to measure the same construct.[11] Support for this interpretation also comes also from a review of education questions in household surveys conducted by the Education Policy and Data Center (EPDC) (2009[22]) for the International Household Survey Network (IHSN). This review recommends the use of single sentence reading test

[10] The question regarding literacy used in earlier waves of MICS and DHS was the following: "Can you read and understand a letter or newspaper easily, with difficulty or not at all?" This question was asked as part of the household questionnaire. The introduction of the single sentence test also meant that the information in literacy was collected from the sampled individual respondent rather than the household head/designated principal respondent.

[11] "Respondents are given a simple test of their ability to read. This test replaces the question in earlier versions of the DHS core questionnaires that asked the respondent for a self-assessment of her literacy" (ORC Macro, 2001, p. 24[69]).

rather than respondent reports as the preferred measure of literacy in household surveys. The absence of any discussion of the differences in the construct measured by the two approaches suggests that the two measures are seen by the authors as measuring the same basic concept.

There are, nevertheless, strong reasons to question the validity of the single sentence reading test as a measure of literacy conceived as the possession of a minimal level of reading comprehension. Most importantly, the single sentence reading test does not attempt to assess *comprehension*. Respondents are asked to read a sentence out loud, not to demonstrate that they *understand its meaning*. From this point of view, it is most appropriately seen as a task assessing an aspect of *decoding skills*, a task in which the objective is to assess the respondent's recognition of familiar words and his/her capacity to sound them out correctly (RTI International, 2015, p. 24[15]). This is the view of commentators such as Schnaffner (2005[23]; 2005[24]) [12] – one of the main proponents of the use of the single sentence reading test – as well as Mingat et al. (2013[25])[13] and UIS (2008[26]).

"Decoding" and "reading comprehension", however, are far from the same thing. While mastery of decoding skills or the understanding of letter-sound correspondences represents a precondition for reading comprehension, comprehension involves far more than word recognition [see RTI International (2015, pp. 23-24[15]) for a discussion]. The other point to note is that decoding itself has a number of dimensions – letter, syllable and word recognition – and that the single sentence test can only be seen, at best, as a very partial test of decoding skills and one which, by virtue of being a one-item test, is subject to considerable measurement error.

What is measured by the STEP literacy core?

The STEP literacy assessment is a version of the assessment used in PIAAC and is underpinned by the same assessment framework [see (Educational Testing Services (ETS), 2014[13]) and (OECD, 2016[17])]. Literacy is defined in terms of the comprehension of written texts, emphasising the role that this plays in effectively acting and engaging in social and economic life. It is defined 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" (OECD, 2016[17])

[12] She describes it as a low level literacy task involving mere "decoding" of written language (Schaffner, 2005, p. 2[23]).

[13] In their words, « [c]e test est certes assez peu exigeant et relève d'une conception restrictive de la lecture, puisqu'il s'agit davantage de déchiffrage que de compréhension…mais il s'avère suffisamment discriminant pour identifier ceux qui ont des difficultés de base par rapport à la lecture ». "This test is not very demanding and is based on a narrow concept of reading as it relates more to decoding than to comprehension…however, it is sufficiently discriminating to identify those who have fundamental difficulties in terms of reading".

and is measured on a continuous scale. *Writing* is not included as part of the construct of literacy and is, therefore, not assessed.

The STEP "core" consists of eight simple literacy items and designed to sort respondents with very low literacy from those with higher levels of skill. The items included in the core assessment are all located at levels 1 or below on the PIAAC literacy scale. The features of items at this level are described in the following terms.

> *The tasks at this level require the respondent to read brief texts on familiar topics to locate a single piece of specific information. There is seldom any competing information in the text and the requested information is identical in form to information in the question or directive. The respondent may be required to locate information in short continuous texts. However, in this case, the information can be located as if the text was non-continuous in format. Only basic vocabulary knowledge is required, and the reader is not required to understand the structure of sentences or paragraphs or make use of other text features (OECD, 2016, p. 71[17]).*

The PIAAC/STEP core can be seen as a test of basic reading comprehension skills. At a minimum, adults who can successfully complete tasks with a difficulty of less than level 1 are able to correctly locate information in a short text. Those adults who fail the core are not necessarily illiterate in the sense that they cannot read words or sentences. However, they lack some or all of the skills needed to effectively understand the meaning of information in text form.

What is measured by the Ghana socio-economic panel test?

The conceptual basis for the reading comprehension test used in the Ghana socio-economic panel (2009[11]) is not explained in the available survey documentation. This assessment task is very similar to reading comprehension tasks administered as part of a short literacy assessment[14] included in a study conducted in Kenya and Tanzania in 1980 (Knight and Sabot, 1990[27]) and reused in the Tanzanian Human Resource Development Survey of 1993 (Ferreira and Griffin, 1996[28]). Little detail is given of the conceptual basis of the assessment in the published reports on these studies. In reporting, the results of the literacy assessment and a short numeracy assessment were combined to form a measure of "cognitive skill".

With the exception of one question (the last, in Box 3), the tasks in this test all involve the location of single piece of information in a short text in which the answer is identical with the terms used in the question. In this sense, the tasks have very similar cognitive demands to those in the STEP/PIAAC literacy core, even if the text is far more scholastic in nature. From this point of view, the test can be seen as an assessment of basic reading comprehension.

[14] That was developed by ETS for the study.

Validity from an empirical perspective

In this section, the empirical evidence regarding the validity of the information on proficiency collected through respondent reports and the different simple direct measures presented above is reviewed.

Respondent reports

Respondent reports regarding literacy proficiency are commonly seen as providing information of questionable validity. They are criticised for being "subjective" (Schaffner, 2005[24]; Education Policy and Data Center, 2009[22]) and, as a consequence, seen as suffering from a range of problems.[15] These include variation between different individuals and groups in the understanding of what it means to be able to read and write, unconscious or conscious misreporting and, in the case of third-party reports, lack of accurate knowledge regarding the skills of the individuals about whom the information is provided.

Respondent reports of literacy proficiency are also commonly claimed to over-estimate true proficiency (OECD/Statistics Canada, 2000[19]; Schaffner, 2005[23]; Schaffner, 2005[24]; Nath, 2007[29]; UNESCO, 2006[30]; UNESCO, 2007[31]; Education Policy and Data Center, 2009[22]; Education Policy and Data Center (EPDC), 2009[32]; Mingat, Ndem and Seurat, 2013[25]). The empirical basis for this claim is, in many cases, rather weak.

Schnaffner (2005[24]), for example, analyses two DHS surveys (in Ethiopia and Nicaragua) in which information on the reading ability of respondents was collected using respondent reports (provided by either the individual concerned or the primary respondent) and the single sentence test. She finds that, in both cases, the proportion of respondents deemed to be able to read on the basis of respondent reports[16] and on the single sentence test increases with years of schooling. However, the proportion of respondents who were reported as being able to read and write was higher than that who could read the single sentence in total and at all levels of schooling with the gap being greatest for those with the least schooling. From this she infers that "subjective literacy measures [i.e. respondent reports] may greatly overstate true literacy rates". It is important to note that this conclusion is "predicated on the *assumption* [emphasis added] that the simple objective measure is a more accurate measure, than the subjective measure" (Schaffner, 2005, p. 656[24]).[17] Unfortunately, no discussion of the basis for this assumption is undertaken, particularly given the strong conclusions drawn from the analysis. It could be legitimately questioned, for example, whether a test consisting of *one item* is capable of providing

[15] Which are by shared by all measures based on respondent reports.

[16] The question was "Does (NAME) know how to read and write, read only, or neither read nor write?"

[17] Leaving aside the issue of the small sample size (two studies) on which Schnaffner's analysis is based, in the absence of any external validation of which of the two measures is closest to the true value, all one can say is that the two approaches give different estimates.

a particularly valid or reliable measure of the capacity of a person to read a simple sentence. A similar conclusion is reached in her background paper for the 2006 EFA monitoring report (Schaffner, 2005[24]) using a larger sample of 11 studies. Here again Schnaffner assumes that the direct assessments provide estimates of proficiency that are closer to the true values than do respondent reports. This is despite the evidence she provides regarding the poor quality of many of the direct assessments she examines which, *prima facie*, should have led the author to be rather cautious in concluding that the test results represent "true" values.

Chapter 7 of the 2006 EFA Global Monitoring Report: *Literacy for Life* restates Schnaffner's conclusions, claiming that "indirect assessments usually overstate 'true' literacy levels" (UNESCO, 2005, p. 180[33]) without offering additional evidence. The 2008 EFA monitoring report (UNESCO, 2007[31]) presents results from the Kenyan Adult Literacy Survey to demonstrate "that conventional data relying on self-assessment tend to overstate actual literacy and numeracy levels" (2007, p. 62[31]). The literacy rates from the direct assessment were lower than the rates calculated in 2000 using data from MICS (based on respondent reports) and also within the Kenyan study, rates of literacy and numeracy based on respondents' reports were higher than those based on the direct assessment.

A review of the Kenyan study suggests that the picture is not quite as clear cut as is presented in UNESCO (2007[31]). In the first place, the literacy assessment appears not to have been administered to respondents who reported that they could not read (28%). This means that the accuracy of self-reports cannot be assessed for this significant group. Second, the extent to which the Kenyan study provides evidence for the claim that respondent reports overstate true literacy levels depends on the how literacy is defined. All the respondents to whom the assessment was administered appear to have at least some basic reading skills such as being able to identify words and understand their meaning (Kenya National Bureau of Statistics, 2007, p. 23[34]). In other words, all persons who took the assessment had at least some decoding skills of the type measured by the single sentence reading test. However, if a more demanding standard is used (that of identifying synonymous matches of words and phrases), the proportion of respondents reporting that they can read is higher than the proportion of those which can correctly undertake this type of task.

Nath (2007[29]) presents the results of a study in Bangladesh that compared reports of literacy with performance on a literacy assessment. In the households included in the study: (1) the household head (or another principal respondent) was asked to indicate whether each of the members of the household aged 11 and above could read and write a letter; and (2) all members of the household aged 11 and above took a literacy test consisting of 24 items. The third-party reports were judged to be accurate in 90.5% of cases. It is, however, difficult to conclude much on the basis of this study. Insufficient information is available to come to an informed judgement regarding the robustness of the conclusions. In particular, very little information is provided either about the test or the criteria used to determine whether a test-taker was assessed to be literate.

Mingat et al. (2013[25]) present the results of an analysis in which they compare illiteracy rates estimated on the basis of self-reports (using data from MICS) and the use of the single sentence tests (DHS) for eight African countries in which the two data sources were available. Their conclusion is the following:

> *En moyenne, les estimations basées sur les enquêtes MICS tendent à proposer des mesures un peu plus élevées du niveau d'alphabétisme des individus que les enquêtes DHS. En effet, 44,9 % des enquêtés de ce programme sont considérés comme analphabètes sur la base de leur déclaration, alors que, lorsqu'ils sont testés, ils sont 48,5 % à ne pas savoir lire (p. 6).*[18]

While this comparison is seen as providing empirical support for the conclusion that "subjective" measures of literacy have a tendency to overstate true literacy rates, there are reasons to be far more cautious. As the authors acknowledge, the differences observed are not large ("*peu marquées*"). No information is provided regarding the size of the standard errors associated with the estimates that would allow evaluating whether they are significant at conventionally accepted levels or not. The major problem with the analysis is that it compares estimates from two different studies using different methods, conducted at different times and under different conditions, not the responses of a single individual to (1) a question about literacy proficiency and (2) his/her performance on a literacy test. As such, it cannot be ruled out that the differences in the estimates relates to the fact that they are derived from different surveys rather than the differences in methods used to derive an estimate of literacy proficiency.

In IALS, in addition to completing a literacy assessment, respondents were also asked how they rated their reading skills in the test language (1) for their main job and (2) needed in daily life (excluding work and school).[19] The response categories were: "excellent", "good", "moderate", "poor" and "no opinion" (Murray, Kirsch and Jenkins, 1998[35]). Responses to these questions are cross-tabulated with actual test performance in OECD/Statistics Canada (1995[36]; 2000[19]). The expected relationships are observed. The proportion of adults reporting "excellent" reading skills "in their main job" and "needed in everyday life" increases with level of literacy and the proportions reporting that they have "moderate" or "poor" declines as does the proportion reporting that they had no opinion or that the question was not applicable [see OECD/Statistics Canada (1995[36]), Tables C-6a-c and C-13a-c]. At the same time, in most countries there are reasonably large proportions of respondents at all levels of literacy who

[18] "On average, the estimates based on the MICS tend to suggest slightly higher levels of literacy than those based on DHS. While 44.9% of respondents to this study are considered as illiterate on the basis of self-reports, 48.5% do not know how to read when tested." [Author's translation.]

[19] There were also questions asking respondents to rate their current reading and writing skills in the first language that they spoke as a child but *only if this language was not the language of the assessment*.

evaluated their reading skills as "excellent" and few who report their skills as "weak" (even at the lowest proficiency level, level 1). The explanation offered for this apparent discrepancy is that respondents may assess the *adequacy* of their reading skills relative to the demands of work and everyday life rather than evaluate the level of their skills *per se* (OECD/Statistics Canada, 1995, p. 127[36]).[20]

Evidence from the British cohort studies is also relevant to the discussion of the validity of respondent reports of own or other's literacy. Both the cohort born in 1958 selected for the National Child Development Study (NCDS) and the cohort born in 1970 participating in the 1970 British Cohort Study (BCS70) were tested for their literacy and numeracy skills and answered questions about difficulties in literacy and numeracy in various waves of these studies. Brynner and Parsons (2006, p. 39[37]) reports the relationship between performance on the literacy tests and self-reported difficulties in reading among the BCS70 cohort with the proportion of respondents reporting difficulties fall as literacy proficiency increases. As in the case of IALS, many respondents with weak skills as measured by the literacy tests, do not report that they have difficulties and vice versa. Echoing the conclusions regarding self-assessments in IALS, the authors comment that:

> *As indicators of skills, such questions reveal discrepancies with the results found from objective tests. Although the two are correlated, many respondents whose test performance is very poor do not acknowledge any difficulty… Similarly, though to a lesser extent, some of those who acknowledge a difficulty have average or better scores on the tests. It seems that self-appraisal is not necessarily grounded in objective evidence of performance but has more to do with self-concept and identity. Do I see myself as poor against the standard that I set for myself in the context of my everyday life? (2006, p. 21[37]).*

In summary, the self- or third-party reports reviewed in the literature discussed above consist of two types: (1) questions focussed on whether the respondent/household member can read or not; and (2) questions focussed on how well the respondent can read.

Studies that compare the reports of whether respondents/household members can read with the results of literacy test are based on a small number of observations and are based on comparisons with "tests" of questionable validity (e.g. the single sentence reading test) or about which limited information is provided (Nath, 2007[29]; Schaffner, 2005[23]).

Regarding the use of self-reports as a means of gaining information regarding how well a person can read, the evidence is stronger. Self-reports

[20] "Although not specifically asked to judge the adequacy of their literacy skill in daily life, this may well have been the criteria that many respondents used. Numerous studies have shown how adults with low literacy skills are able to construct their daily lives so that literacy is not a part of it and therefore, they can legitimately claim that their skills serve them well." (OECD/Statistics Canada, 1995, p. 109[36]).

are unlikely to provide a very valid measure of how well a person can read. The reasons is that this judgement is always going to be a relative assessment against some reference point or benchmark that will vary widely between individuals. In particular, self-evaluations are likely to vary according to the (explicit or implicit) reference point used by the respondent to anchor their evaluation. While this is also true of questions regarding whether a person can or cannot read, the effect is likely to be greater as the reference point is less specific.

Additional analysis using data from STEP

As it contains information regarding reported proficiency in literacy and the results of a literacy, STEP represents a valuable contemporary source of data for comparing respondent reports and direct measures of literacy proficiency. First, all respondents are administered the eight item core reading assessment and an assessment of their reading components skills.[21] Second, two different indirect measures are available.

In waves 1 and 2 of STEP, information was collected on whether respondents who had not completed primary school could (1) read a letter and (2) write a letter in any language as part of the household roster. Response categories were "yes – with difficulty", "yes – without difficulty" and "no". This information was supplied by the *household head* or *designated principal respondent.* This person is not necessarily the same individual as the respondent for the individual sections of the questionnaire and the literacy assessment. In other words, the information on whether the respondent can read or write a letter is in many cases a third-party report rather than a self-report.

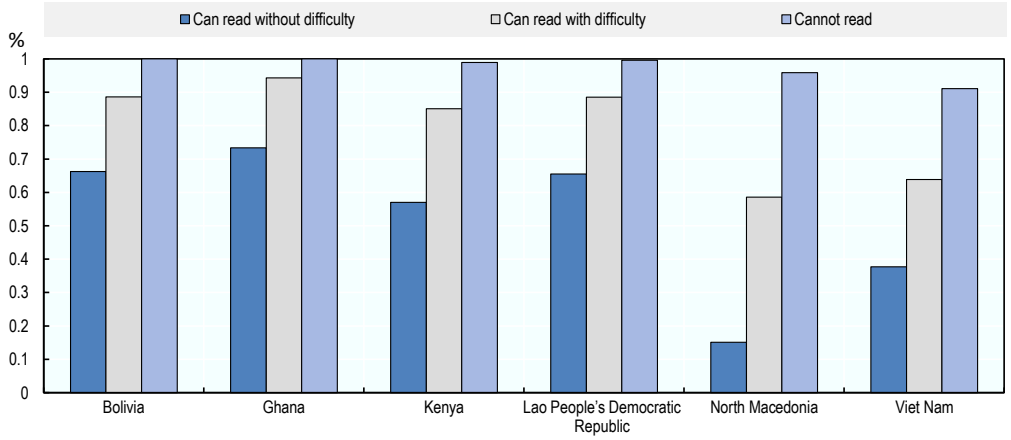
In the second wave of STEP, a question on the languages in which the respondent could read or write well enough to work in a job that requires that language was added. This is answered by *all* sampled respondents. Respondents are presented with a list of languages and asked to indicate whether they can read and write well enough to work in a job that requires that language. Response categories are "yes" and "no".

Among 15-64 year-olds with a highest level of education at primary level, in five of the six countries (the exception being the Republic of North Macedonia) in which there are sufficient respondents with this level of education, most have poor reading comprehension skills in that they fail the literacy core. This is true of almost all those who are reported as being unable to read a letter, most of those who are reported as being able to read a letter with difficulty and a large proportion of those who are reported as being able to read a letter without difficulty (Figure 1).

[21] See note 6 above.

Figure 1. Proportion of respondents failing the core test by reported ability to read a letter

Adults with a highest level of education at ISCED 1 or lower

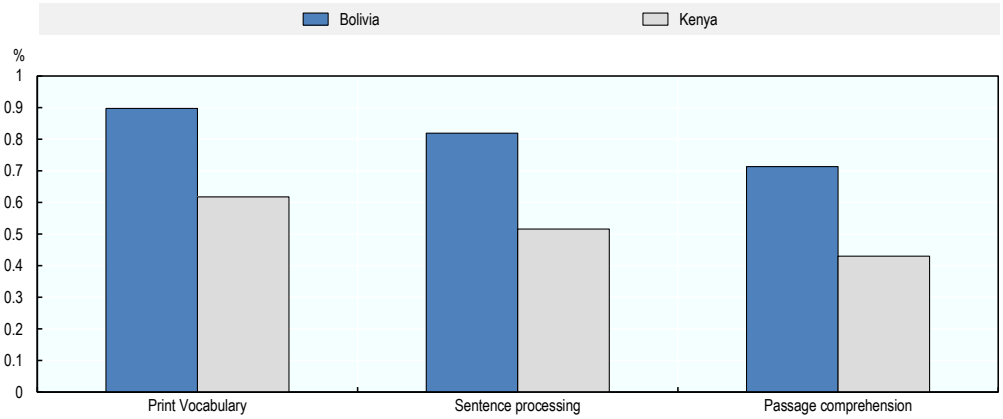


Note: Due to low numbers of respondents in these categories, results for respondents who are reported as being unable to read a letter in the Republic of North Macedonia and Viet Nam, able to read with difficulty in the Republic of North Macedonia and able to read a letter without difficulty in Ghana are not reported.
Source: (World Bank, n.d.[38]), STEP Skills Measurement Household Survey, Wave 1 (2011-13) for Bolivia, Ghana, Lao PDR, Viet Nam and Wave 2 (2012-14) for Kenya and North Macedonia, https://microdata.worldbank.org/index.php/catalog/step (accessed on 6 August 2018).

At first glance, this could be interpreted as indicating a marked tendency to over-estimate reading ability. To examine this question more closely, the performance on the reading components items for respondents who are reported as being able to read a letter without difficulty but failed the core is examined for the two countries (Bolivia and Kenya) in which there are sufficient respondents to examine the question (Figure 2). In Bolivia, many respondents in this group have good vocabulary, sentence processing and passage fluency skills on average. This suggests that it is possible that respondents in Bolivia were reporting their reading skills reasonably accurately, but using a lesser standard for determining whether they could read a letter without difficulty than is implied by passing the STEP core test. In Kenya, however, the performance on the components items of respondents who are reported as being able read a letter without difficulty but failed the core is weak.

Figure 2. Proportion of respondents getting 50% or more items correct in component domains

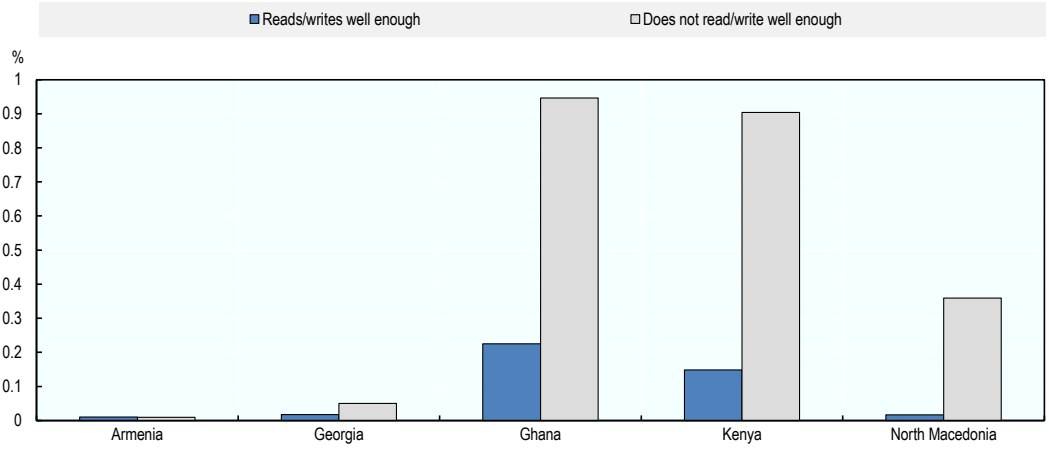
Adults with a highest level of education at ISCED 1 or lower who are reported to be able to read a letter without difficulty but failed the core test, Bolivia and Kenya



Source: (World Bank, n.d.[38]), STEP Skills Measurement Household Survey, Wave 1 (2011-13) for Bolivia and Wave 2 (2012-14) for Kenya, https://microdata.worldbank.org/index.php/catalog/step (accessed on 6 August 2018).

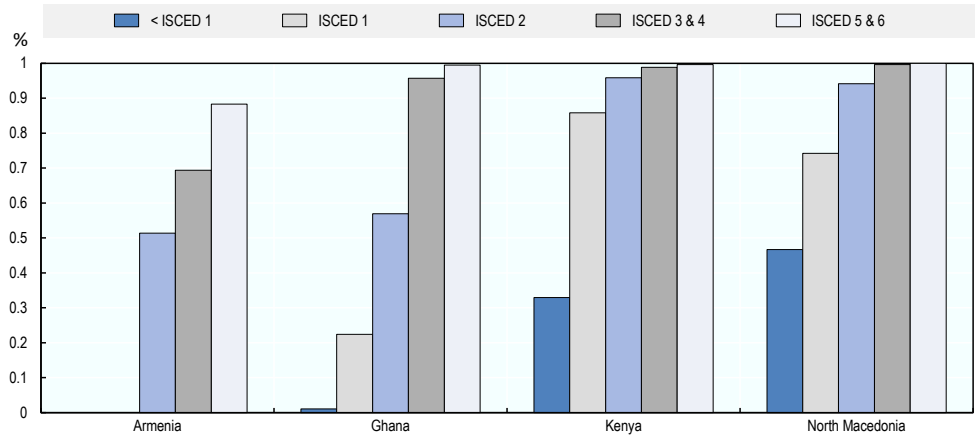
A comparison of respondents' performance on the STEP literacy core with their responses to the questions asking about whether they can read or write the test language well enough to meet the requirements of working in a job in that language suggests that most respondents answer the question reasonably accurately (Figure 3). In the five countries in which data are available, the proportion of respondents who state that they can read and write the test language well enough to work in a job in that language but fail the core is close to zero in three countries and of moderate to small in Ghana (23%) and Kenya (15%). The responses show the expected positive relationship with educational attainment on all countries (Figure 4). In both Ghana and Kenya, among those respondents who state that they can read/write to the standards required in a job but fail the core, most perform reasonably well on the component skill measures (Figure 5). Again, it cannot be ruled out that at least some respondents in this group are giving accurate assessments of their reading/writing ability, particularly if they perceive the reading demands of the labour market as being very low.

Figure 3. Proportion of respondents failing the core test by self-assessed reading/writing skills



Source: (World Bank, n.d.[38]), STEP Skills Measurement Household Survey, Wave 1 (2011-13) for Ghana, and Wave 2 (2012-14) for Armenia, Georgia, Kenya and North Macedonia, https://microdata.worldbank.org/index.php/catalog/step (accessed on 6 August 2018).

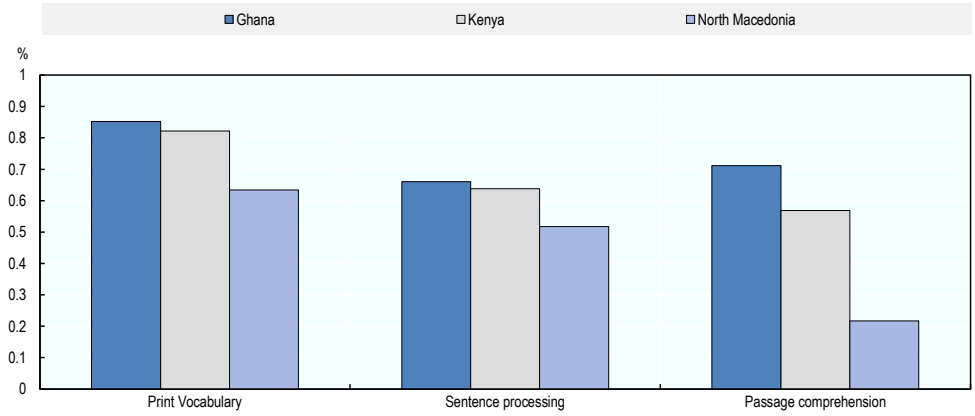
Figure 4. Proportion of adults aged 20-65 years who can read and write well enough in the test language, by level of education



Note: Due to low numbers of respondents, results for adults with level of education at and below ISCED 1 in Armenia are not reported.
Source: (World Bank, n.d.[38]), STEP Skills Measurement Household Survey, Wave 1 (2011-13) for Ghana, and Wave 2 (2012-14) for Armenia, Kenya and North Macedonia, https://microdata.worldbank.org/index.php/catalog/step (accessed on 6 August 2018).

Figure 5. Proportion of components items correct

Respondents who stated that they could read/write well enough in the language if the test to do a job in that language but failed the core test



Source: (World Bank, n.d.[38]), STEP Skills Measurement Household Survey, Wave 1 (2011-13) for Ghana and Wave 2 (2012-14) for Kenya and North Macedonia, https://microdata.worldbank.org/index.php/catalog/step (accessed on 6 August 2018).

In conclusion, at least in the samples examined here, there is some evidence that respondent reports of whether a person can read or not tend to overstate the proficiency of the person concerned as determined on the basis of test performance. However, it is important to acknowledge that the question of the "accuracy" of respondent reports of literacy proficiency can be addressed from two angles. The first is the extent to which they accurately represent the subjective beliefs of the respondent. The second is the extent to which they correspond to an objective or external standard.

In terms of the first question, no evidence is available regarding the criteria that respondents apply to assess their own or others' reading ability. However, the fact that many respondents in STEP who report that they can read well or without difficulty but fail the core literacy assessment have some basic reading skills suggests that they may be reporting their ability against a less demanding set of criteria than is implied by a 'pass' on the core. Differences between self-reported proficiency and "objectively" determined proficiency may be less a result of (unintentional or intentional) error on the part of the respondent than a consequence of the fact that the criteria used by the respondent to judge whether he/she or another person can read a letter or read well in enough in a given language to work in a job that requires that language differ are not well reflected by the "objective" standard used.[22]

[22] In order to establish whether self-reports of reading ability over-estimate real ability, it is important not only to have a reliable and valid direct measure of the ability of the respondent but also some understanding of how respondents interpret the question and of the criteria in terms of which the respondent evaluates his/her own skills and those of others. The fact that a respondent "over-estimates" his/her reading ability or that of others may be due to unintentional or intentional error.

The single sentence test

While the single sentence test is administered only to respondents with a highest level of education of completed primary school or below, it is nevertheless possible to examine the relationship between the probability of 'passing' the test and educational attainment for this group. In the DHS, information is available on the highest grade completed and years of education completed at that level. Table A B.1 in Annex B presents the proportion of respondents who can read the complete sentence in the single sentence test by years of primary education for women aged 15-49 years in 30 Sub-Saharan African countries. In all countries, the association follows the expected pattern with a clear positive association between the number of years of primary school completed and being able to read aloud a simple sentence. This is line with the findings of Schnaffner (2005[23]) for Ethiopia and Nicaragua using DHS surveys of 2000 and 2001 respectively.

There is no published analysis that compares the performance by the same individuals on the single sentence reading test and a test of reading comprehension. The Ghana socio-economic panel survey (2009[11]), which involved the administration of a version of the single sentence test and a short reading comprehension test to respondents aged 9-25 years, appears to represent the unique source of data for examining this question. The comprehension test involved respondents reading a short text on a show card and answering eight multiple choice questions about the text. The text and the questions can be found in Box 2.

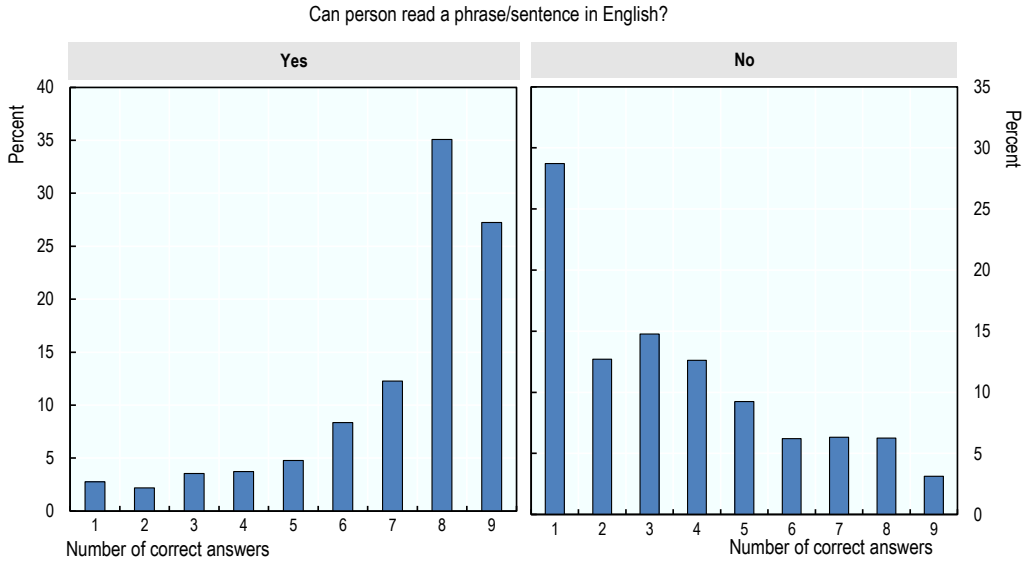
One-thirds (33%) of the 9-25 year-old respondents failed and 64% passed the singe sentence reading test (with 3% of cases missing). Of those failing the sentence test, 46% went on to take the reading assessment. In the case of those who passed, 76% took the reading assessment. The rationale for the selection of young people with a particular result in the single sentence test to take the comprehension test is not clear and it is not possible to know how representative the group that took both tests is of the original sample. The distribution of respondents who did the two tests by number of questions successfully completed in the literacy test according to whether they passed or failed the single sentence test is presented in Figure 6. Overall, those who passed the test performed better on the comprehension test that this who failed. Among respondents who passed the sentence test, very few got any questions incorrect and 88% got at least four questions correct. While 28.9% of respondents who failed the simple sentence test also failed to get any comprehension questions correct, 71% of these

The respondent may share more or less the same understanding of the question as the survey designers but evaluate his/her or other's ability incorrectly due to lack of knowledge or for reasons of social desirability, for example. Alternatively, respondents may not share the same understanding of the question as the designers of the survey and use far less demanding criteria to evaluate whether they can read that than are assumed by survey designers or assumed by analysts. In this latter case, it is more difficult to say that respondent reports 'overstate' real ability. If operationalised as a test score, the respondent's concept of what it is to read may be less demanding than that assumed by the test designer.

respondents got at least one question correct and 30% got at least four correct.

Figure 6. Number of items correct in the reading comprehension test by whether respondent passed single sentence test

Ghana socio-economic panel, 2009



Source: Adapted from Institute of Statistical Social and Economic Research - University of Ghana (2009[11]), *Ghana - Socio-economic Panel Survey,* http://microdata.worldbank.org/index.php/catalog/2534.

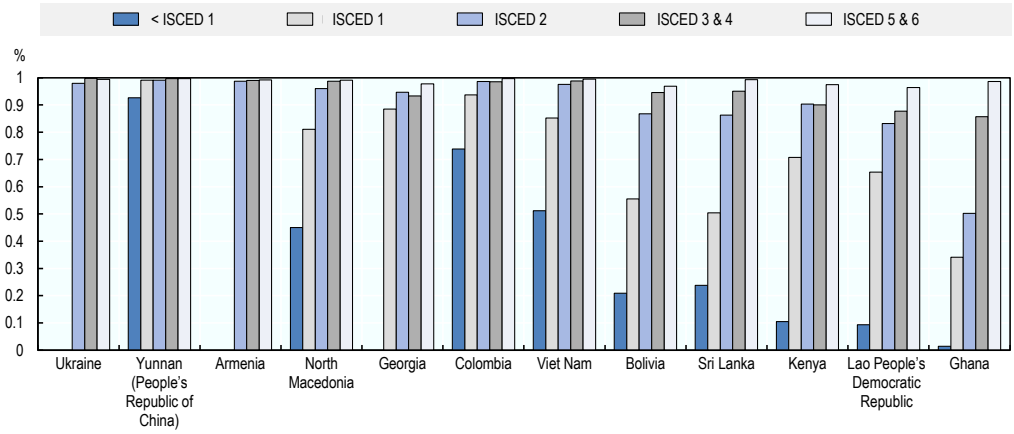
It would be tempting to conclude that the risk of error is greater for those who are assessed as failing the sentence test than for those who pass – i.e. that the test understates reading ability. However, the fact data are available for only one country and that less than half the respondents who failed the sentence reading test and a quarter of those who passed did not take the comprehension test urges some caution in drawing any strong conclusions from these data.

The STEP core as a measure of basic reading comprehension

To examine the validity of the STEP core test as a measure of basic reading comprehension, two main indicators are examined: the relationship between educational attainment and performance on the core and performance on the core and performance on reading components tasks.

Given the role of schooling in the development of literacy, it would be expected that adults who had not attended school or who had very low levels of schooling would have a lower probability of passing the core than other adults. Figure 7 shows the proportion of respondents passing the STEP literacy core by level of highest education for 12 countries participating in STEP.

Figure 7. Proportion of respondents passing the STEP core, by highest level of education



Note: Due to low numbers of respondents, results for respondents with level of education at and below ISCED 1 in Armenia, Georgia and Ukraine are not reported.
Countries are ranked in descending order of the percentage of respondents passing the core test.
Source: (World Bank, n.d.[38]), STEP Skills Measurement Household Survey, Wave 1 (2011-13) for Bolivia, Colombia, Ghana, Lao PDR, Sri Lanka, Ukraine, Viet Nam and the Yunnan province in China, and Wave 2 (2012-14) for Armenia, Georgia, Kenya and North Macedonia, https://microdata.worldbank.org/index.php/catalog/step (accessed on 6 August 2018).

The relationship between the proportion of respondents passing the core assessment and their level of educational attainment is mediated by a number of factors. First in all the countries in the data set, adults with a full secondary education or higher have very high chances of passing the core. This is also true for respondents with lower secondary education in many though by no means all countries. Second, in some countries there are negligible numbers of respondents with a highest level of education at less than primary level (Armenia; Georgia; Ukraine; Yunnan, China) or even primary level (Armenia, Ukraine). In countries where there are largish proportions of adults with very low levels of education, the chances of passing the core are lowest for adults who have not completed primary school than for all other groups followed by adults who have a highest level of education at primary level.

All respondents in STEP undertake a test of reading components as well as the core literacy assessment. The literacy components assessment consists of an assessment of three domains of knowledge and skills that are seen as essential preconditions for effective reading comprehension: print vocabulary, sentence processing and passage fluency. Table 2 presents the average proportion of items correct in the three domains of the components assessment for respondents who passed and for those who failed the literacy core. Only those countries in which at least 4% of respondents failed the core are included.[23]

[23] Armenia (99%), the Republic of North Macedonia (98%), Ukraine (99%) and Yunnan (People's Republic of China) (99%) are not included.

Table 2. Proportion of items correct in reading components by performance on literacy core, STEP

		Proportion correct in Print Vocabulary		Proportion correct in Sentence Processing		Proportion correct in Passage Comprehension	
	Proportion passing core	Passed core	Failed core	Passed core	Failed core	Passed core	Failed core
Bolivia	0.84	0.99	0.81	0.9	0.66	0.95	0.61
Colombia	0.96	0.98	0.77	0.91	0.65	0.94	0.55
Georgia	0.96	0.97	0.32	0.93	0.26	0.93	0.22
Ghana	0.50	0.94	0.35	0.82	0.23	0.91	0.24
Kenya	0.76	0.97	0.56	0.88	0.39	0.92	0.38
Lao People's Democratic Republic	0.54	0.96	0.37	0.82	0.25	0.85	0.15
Sri Lanka	0.81	0.97	0.71	0.86	0.47	0.87	0.21
Viet Nam	0.95	0.98	0.71	0.89	0.56	0.93	0.43

Source: (World Bank, n.d.[38]), STEP Skills Measurement Household Survey, Wave 1 (2011-13) for Bolivia, Colombia, Ghana, Lao PDR, Sri Lanka, Viet Nam and Wave 2 (2012-14) for Georgia and Kenya, https://microdata.worldbank.org/index.php/catalog/step (accessed on 6 August 2018).

Adults who pass the STEP literacy core have a very high probability of getting almost all items correct in all three reading component domains. In contrast, adults who fail the core generally get a considerably lower proportion of items correct. In addition, the time taken to complete the set of items in each of the domains is generally considerable longer for those adults who fail the core than for those who pass.

Among the OECD countries participating in PIAAC, broadly similar results are found. Respondents passing the core have very high chances of getting all items correct. Respondents failing the core get a lesser proportion of items correct. However, the gap in performance on the components items is less in the case of OECD countries in PIAAC than for the countries in STEP. What differentiates respondents passing from those that fail is the average time taken to complete the components assessment is much greater for those who fail than for those who pass (between 1.6 and 1.8 times greater depending on the domain).

In summary, there is empirical evidence in support of the interpretation of a "pass" on the PIAAC/STEP literacy core as an indicator of basic reading comprehension skills. Adults who fail the core are not necessarily "illiterate" in the sense that they cannot read words or phrases, for example. However, even if they have some vocabulary knowledge and knowledge of the semantic knowledge of sentences, they have not mastered this at the level of automaticity necessary for effective comprehension, even at a very minimal level.

Summary: Validity

The review of the validity of respondent reports and simple direct measures if literacy proficiency undertaken above concentrated on: (1) construct validity or the extent to which the measures reviewed are designed to measure the concept of interest – in this case literacy understood as a basic

level of reading comprehension; and (2) the extent to which there is some empirical evidence of the validity of these measures.

Regarding construct validity, two clear conclusions can be drawn. The first is that there is very little information provided by most studies about the concept of literacy that they are trying to measure. The exception is the World Bank's STEP measurement study. The second is that the most commonly used simple direct measure of literacy – the single sentence test – should not be seen as a measure of the construct of primary interest in the measurement of literacy, that of reading *comprehension*. It is most appropriately seen as a test of a component of decoding skills.

The conclusions regarding validity from an empirical perspective are summarised below.

Respondent reports

Respondent reports are commonly presented as providing poor quality information about literacy proficiency. In this paper, a distinction is made between respondent reports of whether a person can read and reports about how well a person can read. In respect of the question of whether a person can read, STEP represents the one study in which it possible to compare the results of questions regarding the ability to read with the results of well validated tests of reading ability. The evidence points to a tendency for respondents to over-estimate their own or others' ability to read. The magnitude of any effect of this type varies according to the "objective" reference point chose for comparison (e.g. passing the STEP core or proportion of components items correct). It also varies by country.

Regarding the question of how well can a person read, the evidence from IALS and the British Cohort Study (BCS) suggests that self-reports of proficiency are not particularly strongly correlated with the results of tests of reading skills. The assessments respondents make are always relative to a subjectively determined point of reference. The reference points of contexts against which or in terms of which adults assess their own proficiency are likely to vary enormously between respondents.

The single sentence test

The performance of respondents on the single sentence reading test has the expected positive relationship with years of education. There is, however, no evidence concerning the relationship between performance on this 'test' and a well validated reading assessment. The evidence from the Ghana Socio-Economic panel of 2009 (2009[11])is insufficient to draw any conclusions.

The single sentence test and respondent reports both represent imperfect measures of whether a person can read. Given the data available, it is difficult to determine which of the two is the most imperfect.

The STEP core

There is strong evidence from both STEP and PIAAC that the PIAAC/STEP literacy core test represents a valid measure of whether or

not an adult has a basic level of reading comprehension understood being able to correctly locate information in a short text. In particular, the STEP core appears to discriminate well between individuals with weak and strong reading component skills.

Comparability: Respondent reports of literacy proficiency

As can be seen from Table A A.1, there is considerable variation in the questions used to collect information on literacy proficiency in surveys as well as in the response categories used and in the exact populations covered by questions. This has been noted previously by the EPDC (2009[32]; 2009[22]). The purpose of this section is to give a systematic description of the differences between studies in terms of:

- The content of questions regarding proficiency
- Response categories used
- Populations covered
- Whether information is collected in the form of self- or third-party reports.

Content

In examining differences in the content of questions regarding literacy proficiency, the focus is on the extent and the manner in which literacy is specified in survey questions. Two broad types of questions can be identified: the first focusses in whether the person of interest (respondent or household member) is literate (i.e. can read or write), the second focusses on how well the person of interest can read (i.e. on performance or proficiency).

The question about whether a person (the respondent or a household member) is literate involves first a decision regarding what dimensions of literacy are covered (e.g. reading and writing) and, second, decisions about whether the question about whether a person is literate is asked as general question (e.g. "Can [NAME] read?") or be specified in some way, in the sense that the act of reading or writing is defined in terms of additional characteristics or attributes. The extent of additional specification varies. Four broad types of specification of the act of reading and writing are found: specification of reading and writing in terms of (1) the reading (and writing) of specific types of documents, texts or elements of text; (2) languages, (3) performance criteria; and (4) context.

Broad dimensions

Surveys vary in the extent to which they operationalise the concept of literacy as one of literacy, of reading or of reading and writing. The following options are found in the surveys reviewed:

- Is [NAME] literate?
- Can [NAME] read?

- Can [NAME] read? AND Can [NAME] write? (two separate questions)
- Can [NAME] read AND write?
- Can [NAME] read OR write?
- How do you rate your reading skills?
- How do you rate your writing skills?

Text types

In many, though by no means all surveys, the questions about reading and writing ask about the reading and writing of particular types of written material, documents, texts or elements of written language (e.g. "Can [NAME] read *a certain type of text*?"). When reading and writing is defined in terms of as the reading/writing of defined types of text or text components, there is considerable variation between surveys. Table 3 provides examples of the types of documents, texts and text components in terms of which reading and writing are defined in different surveys.

Table 3. Specification, by text type: Reading and writing

Can [NAME] read…	Can [NAME] write…
No document type(s) specified	No document type(s) specified
A newspaper	A letter
A letter	A one page personal letter
A simple letter	A short simple statement
A short simple statement	A phrase
Road signs	His/her name
Books…	Fill in forms

Language

The act of reading and writing is also often specified in terms of the language(s) in which a person reads or writes. The approaches adopted regarding the specification of the languages in which a person is literate or can read and/or write are the following:

- There is no mention of language (e.g. "Can [NAME] read?")
- The specification is in terms of all languages (e.g. "Can [NAME] read *in any language*?")
- A limited number of languages are specified (e.g. "Can [NAME] read in *language x*?" and "Can [NAME] read in *language y*?")
- A limited number of languages are specified with the addition of an "other" category (e.g. "Can [NAME] read in *language x*?" and "Can [NAME] read in *any other language*?" Alternatively: "What is your level of reading on your mother tongue?", "What is your level of reading in your second/third… language?")
- The respondent is asked to specify the languages in which he/she can read or write. This can be limited to a particular group of

languages (e.g. "In which of the languages *of country x* can [NAME] read?") or left completely open (e.g. "In which languages can [NAME] read?").

Performance criteria

Further specification of the level or type of reading or writing is observed in the studies examined in this paper in terms of the level of performance demanded of the reader from a cognitive or functional perspective. An example of specification in terms of cognitive performance is the following: "Can [NAME] read in the English language *with understanding*?" An example of specification in terms of functional performance (in this case, defined in relation to the demands of the labour market) is: "In which languages does [NAME] read and write *well enough to work in a job that requires that language*?"

Context

The context in which the activity of reading or writing take place can also be specified. An example of this is: "How would you rate your reading skills in [language x] *for your main job*?" and "How would you rate your reading skills in [language x] *needed in everyday life*?"

Response categories

There is some variation across surveys in the response categories used for similar questions with some studies using a two point and others a three-point scale. In the case of questions concerning the capacity to read and/or write, the response categories are generally "yes" or "no" or a variation of the following: (1) cannot read or write, (2) can read only, and (3) can read and write.[24] When questions are asked about reading and writing separately, the response categories are again either "yes" or "no" or a variation of (1) Yes, easily, (2) Yes, with difficulty and (3) No.[25] In both cases, the two response scales are compatible in that the three-point scale can be recoded into the two point (yes/no) scale.

Populations covered by questions on literacy-

Leaving aside the different target populations of the studies concerned,[26] studies differ in terms of whether or not information on literacy is collected about or from the all in scope individuals or respondents or a subgroup of in scope individuals and respondents. Age and educational attainment are the two variables which serve to define the sub-populations of respondents about which or from whom information on literacy is collected.

[24] The other logically possible option of being able to write only is never available, presumably on the grounds that it is impossible to write without being able to read.

[25] A four-point response scale is used in the South African General Household Survey (2015[55]).

[26] For example (for individual interviews): adults aged 15-65 years in STEP, women aged 15-49 years in MICS and adults aged 15-49 years in DHS.

Age

When information on literacy is collected as part of a household questionnaire (i.e. a questionnaire collecting information about all members of the household), a minimum age threshold is always defined. In the sample of studies used in this paper, the collection of information on proficiency (whether individuals can read and/or write), there are examples in which the minimum age of persons about whom information is collected in respect of household members starting the ages of is 3, 4, 5, 6 and 10 years of age.

Educational attainment

Many studies restrict the collection of information about literacy proficiency to respondents with low levels of educational attainment. Among the studies that filter the collection of information in this way, there are differences in the attainment threshold applied. Examples include:

- Household members who have *never attended school*.
- Household members with a highest level of attainment at *less than primary.*
- Household members with a highest level of attainment *at primary school and below.*

The attainment filters applying may differ within the same study. Wave 2 of STEP is a case in point. In the household component of the questionnaire, information is collected about whether household members with a level of educational attainment at *primary school or lower* can read and write. In the individual component *all respondents* (irrespective of their level of attainment) are asked about the languages in which languages they read and write well enough to work in a job that requires that language and *all respondents* take the reading components and core literacy assessment.

Third-party or self-reports

Surveys differ as to whether they collect information about reported literacy proficiency as part of a household questionnaire, an individual questionnaire or both. In a number of studies reviewed, information about literacy proficiency is collected through a household questionnaire for which the respondent is the household head or designated principal respondent. In this case, the respondent reports information about him/herself as well as all other in scope household members. This is the case for census collections (often) and household studies in which the principal unit of interest is the household. In other studies, information about literacy is collected in individual questionnaires in which the respondent provides information about him/herself alone. Some studies such as STEP Wave 2 collect information about reported literacy proficiency as part of the household roster and also in the individual questionnaire.

Reading and writing

When information is collected on both reading and writing using separate questions, in a number of surveys (though not all), the question on writing is only asked about or to individuals who can read.

The correlation of reading and writing

The UN definition of literacy cited above emphasises both the capacity to read and write a short statement. As has already been noted, the questions that are used to measure literacy in household surveys and censuses vary widely in terms of: (1) whether they ask about reading *and* writing; and (2) when information is sought about both, the form in which this is done.

An important question is whether there is additional value in asking questions about both reading and writing either as a single question ("Can [NAME] read *and* write?") or as separate questions ("Can [NAME] read?" and "Can [NAME] write?"). If there is a high degree of correlation between respondents' reports of their capacity to read and to write, there is little reason to ask about writing in addition to reading.

A comparison of the responses to the questions regarding whether household members with a highest level of education at primary level or less can read and can write in STEP shows a high degree of correlation between answers to the two questions. As can be seen in the four countries covered, in 5% or less of cases, it is reported that a person can read but not write or vice versa (Table 4).

Table 4. Proportion of respondents reported to read but not write or write but not read, STEP

Bolivia	Ghana	Lao People's Democratic Republic	Sri Lanka
0.05	0.02	0.01	0.05

Source: (World Bank, n.d.[38]), STEP Skills Measurement Household Survey, Wave 1 (2011-13), https://microdata.worldbank.org/index.php/catalog/step (accessed on 6 August 2018).

Educational attainment as a proxy for literacy

As has been noted above, in a number of studies, information regarding literacy proficiency is collected from a subset of the population as it is assumed that adults with a given minimum level of completed education are literate. Two comments are relevant here. First, studies vary in terms of: (1) whether or not they restrict the collection of information regarding literacy depending on the level of education of respondents/household members; and (2) in the attainment threshold which applies. This affects comparability between studies.

Second and more importantly, the failure to collect information from or about individuals with relatively low levels of educational attainment (e.g. ISCED 2 or even 3) in some countries will lead to a considerable over-estimation of the literacy rate for the whole population. This is

especially true if the threshold is set quite low, for example, at primary level. As can be seen from the data from STEP on the relationship between passing the STEP core and level of educational attainment, in some countries there is a sizable proportion of the population who have completed lower secondary education who could be considered as lacking basic reading comprehension skills.

To illustrate this point, estimated literacy rates (using passing the core test as the indicator of literacy) for 12 STEP samples for which data are available have been calculated using different assumptions regarding the literacy rate of adults with different levels of attainment and compared with the actual proportion of respondents passing the core. The estimated rates are calculated as the proportion of adults: (1) passing the core or having a level of education of ISCED 4 or higher; (2) passing the core or having a level of education of ISCED 3 or higher; and (3) passing the core or having a level of education of ISCED 2 or higher (Table 5).

Table 5. Estimated literacy rates: effects of assuming 100% literacy at different levels of educational attainment

	Proportion Passing core	Passing core or ISCED4+	Passing core or ISCED3+	Passing core or ISCED2+
Armenia	0.99	0.99	1.00	1.00
Bolivia	0.84	0.85	0.87	0.90
Colombia	0.96	0.96	0.96	0.96
Georgia	0.96	0.97	0.99	1.00
Ghana	0.50	0.51	0.54	0.70
Kenya	0.76	0.76	0.80	0.82
Lao People's Democratic Republic	0.54	0.54	0.56	0.59
North Macedonia	0.98	0.98	0.99	0.99
Sri Lanka	0.81	0.81	0.83	0.88
Ukraine	0.99	1.00	1.00	1.00
Viet Nam	0.95	0.95	0.96	0.96
Yunnan (People's Republic of China)	0.99	0.99	0.99	1.00

Source: (World Bank, n.d.[38]), STEP Skills Measurement Household Survey, Wave 1 (2011-13) for Bolivia, Colombia, Ghana, Lao PDR, Sri Lanka, Ukraine, Viet Nam and the Yunnan province in China, and Wave 2 (2012-14) for Armenia, Georgia, Kenya and North Macedonia, https://microdata.worldbank.org/index.php/catalog/step (accessed on 6 August 2018).

As can be seen, assuming that adults who have completed less than a full secondary education are literate is likely to have a significant impact on estimated literacy rates in some countries. For example, in Ghana, the estimated literacy rate is 70% when it is assumed that all adults with ISCED 2 and higher attainment are literate, some 20 percentage points higher than the "true" rate in this scenario. Assuming that all adults with a completed secondary education are literate also leads to an over-estimate in literacy rates though of a lesser magnitude in some countries. While this exercise has value only for the purposes of demonstration, it points to the need to carefully review any educational attainment filters that apply to questions about literacy proficiency or to the administration of direct measures, particularly in low-income countries. For many of the countries

covered by the MICs and DHS programmes, for example, the fact that respondents with less than a full secondary education do not have their literacy proficiency tested is likely to result in a significant over-estimate of literacy rates.

Summary and discussion

There is considerable variation between studies in the collection of information on literacy proficiency through respondent reports. This covers the broad concept of literacy measured, the extent of specification of the concept, response categories, the populations about which information is collected and the extent to which information is in the form of self- or third-party reports. This variation has an inevitable effect on the comparability of data between studies. Differences in the extent of the specification of reading or writing in terms of text types and language, for example, mean that different studies may collect information about quite different concepts of literacy. Being able to reading a letter is not the same thing as being able to read a newspaper or being able to read in general. Similarly, being able to read in the national language or languages is not the same as being able to read in any language. Considerable differences between the questions used to elicit information about basic literacy are observed within the same survey programmes in some cases. For example, in surveys conducted as part of the World Bank's LSMS programme in Malawi, Niger, Nigeria, Tanzania and Uganda between 2010 and 2016, each survey used a different question regarding literacy. Differences in the exact questions asked may have a significant impact on the extent to which results from different studies can be directly compared. Noting the 10 percentage point difference in estimates of the literacy rate derived for two studies conducted in 1998 and 2000[27] in Côte d'Ivoire, République de Côte d'Ivoire, (2000, p. 29[39]) comments that *"[u]n tel écart entre les deux opérations pourrait provenir d'une différence de definitions"*.[28]

The fact that studies vary in terms of whether information is collected from the household head or designated principal respondent in the form of third-party reports or from a sampled individual in the form of self-reports, may also have an effect on comparability. However, the available evidence regarding differences between the responses provided in the form of self-reports or of third-party (or proxy) reports is rather mixed. In a review of the literature that is now admittedly rather old, Moore (1988[40]) finds "little support for the notion that self-response survey reports are of generally better quality than proxy reports".

[27] The 1998 study (RGPH980) defined literacy as being able to read or write in any written language whereas the 2000 study (MICS) defined literacy as being able to read a letter or a newspaper.

[28] "A gap of this size between the two studies may be due to a difference in definitions."

5. Conclusion and recommendations

The purpose of this paper has been to undertake a review of the indirect and simple direct measures of literacy commonly used in household surveys such as census and omnibus surveys such as health, demographic and living standards surveys. The main findings are the following:

- Insufficient attention is given to the definition of the construct or concept of literacy that is intended to be measured in most of the survey programmes reviewed.
- Both respondent reports and the most commonly used simple direct measure of literacy (the single sentence test) represent imperfect measures of the ability of an adult to read (or write).
- Respondent reports do not provide valid information regarding how well a person can read (or write).
- The example of STEP and PIAAC demonstrates that it is possible to have a test consisting of a small number of items that will provide reasonably accurate evaluation of whether an individual has basic reading comprehension skills or not.
- In many low-income countries, care should be taken in using educational attainment as a proxy for the capacity to read. It should not be assumed that all adults who have completed primary schooling or who have attended some secondary level education can read.
- There is considerable variation the way in which questions on reading and writing are specified and formulated as well as in the populations to which they are asked across household surveys and population census. While many other factors affect comparability, variation in these dimensions will inevitably have an impact on the comparability of literacy statistics from these sources.

The above suggests that there is considerable scope for the improvement of the measurement of literacy (and numeracy) using indirect and simple direct measures of literacy in household surveys. Given the need for and interest in having good information on literacy and the fact that census and omnibus household surveys are likely to remain the principal source of information on the literacy proficiency of adults in many low- and middle-income countries, action to improve the quality of this information should be a priority for the international community and agencies such as UNESCO that have a mandate in this area. In this spirit, some recommendations for the improvement of the collection of information on literacy in household surveys are presented below.

The recommendations focus on measures to improve the validity and comparability of the measures of literacy proficiency collected in household-based surveys and population censuses. There are many other potential sources of error in survey-based research that are also relevant to the collection of information on literacy, in particular, “representation errors” (Groves and Lyberg, 2010[41]) or errors associated with the

inferences from respondents to a target population (coverage errors, sampling errors and non-response). (United Nations, 2005[42]) and Survey Research Center (2016[43]), for example, provide comprehensive guidance relevant to the design and implementation of household surveys in low- and middle-income countries.

Develop a well-defined and shared concept of literacy

The first step in developing good measures of literacy is to have a clear concept of what it is that is being measured. As has been noted, with the exception of STEP/PIAAC, there is very little information available regarding the construct of literacy that is being measured in the studies reviewed in this paper. The apparent failure to pay more attention to the definition and description of the construct of literacy intended to be measured can be seen as contributing to the variation in the measures used and also to the use of the single sentence reading test as a measure of literacy in the MICS and DHS programmes. In the case of the latter, it seems that the choice to use the single sentence reading test may have been made with little examination of the extent to which a measure of decoding would constitute an acceptable proxy for reading comprehension.

From this point of view, there is a need to develop a shared understanding of the construct of literacy that should be measured, particularly the conceptualisation of the threshold between pre-reading skills and reading for meaning. The PIAAC frameworks offer a good starting point. This should be one of the main tasks to be undertaken as part of the development of international guidelines regarding the collection of information regarding literacy in household surveys recommended below.

Where possible collect information on literacy through valid and reliable tests

The collection of information on the literacy proficiency of the adult population can be seen as focussing on two broad questions: (1) what proportion of the population is literate; and (2) what is the distribution of proficiency among the literate population. The preferred methods for the collection of information on these two questions are detailed in Table 6.

Table 6. Hierarchy of preferred measures of literacy

	Above/below comprehension threshold	Reading proficiency
Preferred method	Assessment (short)	Assessment
Fall-back	Respondent reports	Estimation (in a limited set of circumstances)

Two comments are in order. First, the preference for tests as a means of gathering information on literacy should not be seen as a blanket endorsement of tests over other methods. If tests are to be used to assess literacy, they should meet accepted standards of validity and reliability. From this point of view, it would not be recommended that the single

sentence test be used in place of respondent reports.[29] Second, regarding information on the distribution of proficiency in literacy (i.e. how well adults read), respondent reports do not represent a viable fall-back option.

Develop a short assessment of basic reading comprehension for use in omnibus household surveys

The most commonly used short "test" of literacy used in household surveys is the single sentence reading test administered by MICS and DHS. This cannot be seen as a test of basic reading comprehension. It is also likely to be have limited validity and reliability as a test of reading aloud or decoding skills simply because it is a one-item test.

The use of a more construct-relevant measure would also have considerable analytic benefits more generally. A better indicator of basic comprehension skills would have value in health surveys for instance, particularly in the context of understanding the capacity of adults to effectively deal with information about health in written form. A measure that focussed on comprehension would improve understanding about the likely take-up of such information.

The core literacy assessments of STEP and PIAAC show that it is possible to have a valid and reliable measure of a minimum level of proficiency in reading comprehension based on a test of eight items. It would be possible to develop a set of similar items designed to measure whether a respondent was above or below the basic comprehension cut-off defined as the ability to read brief texts on familiar topics to locate a single piece of specific information. These tasks could be designed to be answered orally.

It is recommended that the DHS and MICS programmes develop a short reading test along the lines suggested above to replace the single sentence reading test that is currently used. The objective would be to develop a set of items that could be administered in a relatively short time focussed on the basic reading comprehension threshold that would have a strong theoretical basis in a validated assessment framework such as that of PIAAC.[30] These items should be designed for use in low-income countries. As is done in the PIAAC and STEP core tests, successful completion of a minimum number of items would be set as the indicator of whether the respondent was above or below the comprehension threshold.

An important question that would need to be addressed in the process is whether such test should be linked to PIAAC as the only existing international comparative assessment of literacy in the field at the time of writing. A link to PIAAC would provide results that could be interpreted in terms of a validated scale. However, establishing a psychometric link

[29] At the same time, this should not be interpreted as suggesting that that MICS and DHS return to the use of respondent reports rather than the single sentence test. Improvement in measurement of literacy in MICS and DHS can only come through the development and use of a more valid and reliable short test.

[30] This can be seen as a version of the proposal developed by Gal (2016[10]) but with far more modest aims.

would entail the conduct of linking studies in which the new items together with PIAAC items were administered to the same individuals. The alternative would be to seek to achieve comparability within survey programmes – i.e. that successfully completing the defined minimum number of items reflected the same minimum level of proficiency across the participating countries – rather than with the PIAAC scale.

Standardise the collection of information on literacy through respondent reports

There is considerable variation in the exact questions used in different surveys concerning the ability of respondents and household members to read or not. There would be value in developing a standard recommended approach to the wording and response categories of such questions in population censuses and household surveys. In low-income countries with a large number of languages, asking people whether they can read (and write) may represent the most feasible and cost effective way to collect data on basic literacy. To the extent that respondent reports as to whether a person can read or not are likely to continue to represent an important source of information on literacy in low- and middle-income countries, every effort should be made to ensure that data from these sources are as comparable as possible. While lack of comparability arises from many sources, variation in the definition of the concept measured and in the wording of questions is one that is potentially significant which could be eliminated relatively easily.

Review the educational attainment thresholds applying to questions regarding literacy proficiency and the administration of simple direct measures of literacy

As noted above, many studies administer questions about literacy proficiency or simple direct measures of literacy to respondents with very low levels of educational attainment only. The reasons for this are understandable in terms of reducing respondent burden and maximising co-operation, especially for groups such as people with university level attainment, in which the members are likely to have rates of literacy approaching 100%. However, in low-income countries and countries with multiple languages and high levels of immigration, there are reasons to question or test as much of the population as possible. The evidence from STEP is that many studies use too far low a cut-off point. In particular, there is likely to be a reasonable proportion of adults who report having completed some lower secondary schooling as well as some adults who have completed higher levels who lack basic reading comprehension skills in many low-income countries. Assuming that all persons with lower secondary or even upper secondary level education are literate will result in an over-estimate of literacy rates. In some countries, the impact is likely to be substantial.

Do not collect information about writing in addition to reading skills

Conceived broadly, literacy encompasses both the ability to read and to write. This is reflected in the UN definition cited above. The two skills are highly related, in that to write one must be able to read (even if the reverse is not true). Empirically, the correlation between answers to questions about whether individuals can read and can write is extremely high. When the two questions are asked independently, almost all adults who indicate that they can read also indicate that they can write. At least in terms of *self-reported* information, there seems little reason to ask about both reading and writing skills. The bigger question is whether writing skills should be tested through some simple direct measure. The answer is "no", given the current state of development of the assessment of literacy among adults. First, a very high correlation between basic comprehension and basic writing skills would be expected. Second, and perhaps more importantly, a well-developed conceptual framework for understanding writing skills and for assessing them among adults does not exist.

Develop international guidelines regarding the collection of information regarding literacy in household surveys

In reviewing the measures of literacy used in omnibus household studies, two facts stand out. First, there is a lack of documentation of the measures of literacy implemented in household surveys. Second, the considerable variation in the questions used to collect information in literacy. These two facts are to some extent related. Overall, the absence of a well-articulated clear and broadly shared understanding of the concepts to be measured increases the risk that poor measures are used and that there is variation between studies.

For most of the survey programmes and individual surveys reviewed in this paper, the available documentation (where it exists) contains very little information regarding the measures of literacy used. The exception is the World Bank's STEP measurement study in which the measurement of literacy is very well documented (Educational Testing Services (ETS), 2014[13]; Gaëlle et al., 2014[5]). The limited nature of the documentation regarding the measurement of literacy in other studies is understandable. In omnibus studies such as the DHS and MICS, where the focus is on the collection of information on a wide range of health-related topics, literacy represents just one of the aspects of the individual respondent's background about which information is collected.

At the international level there is also a lack of any comprehensive and up-to-date guidance for the measurement of literacy. The measurement of literacy is discussed in the UN *Principles and Recommendations for Population and Housing Censuses* (United Nations (UN), 2015[14]). However, it is covered in one paragraph. A useful and comprehensive discussion of the measurement of literacy in household surveys was published in 1989 by the United Nations Department of Technical

Co-Operation for Development and Statistical Office (1989[44]).[31] While much of what is discussed in this document remains relevant today, there have been considerable developments in the measurement of literacy in an international comparative context over the last 30 years.[32] A proposal for a literacy module in household surveys was developed in UNESCO (2008[45]). Measures of literacy are also discussed in EPDC (2009[22]) and (2009[32]) in the context of the measurement of education in household surveys, though the focus of these papers is less on the concepts than data items and survey design.

There would be considerable value in developing an international framework and guidelines for the measurement of literacy that could serve as the basis for the greater harmonisation of the collection of information on literacy. This should cover the collection of information on:

- Proficiency
- Literacy and literacy-related practices
- Participation in literacy programmes
- The literate environment.

Work on harmonisation should take place at both the level of the definition of constructs and concepts and that of measures. As a starting point, there is a need to provide a conceptual framework to guide the collection of information regarding the collection of information on literacy in household-based studies. This should identify:

- The constructs and concepts that should be measured in the relevant domains, together with the rationale for measuring them.
- The definitions of the concepts and constructs proposed to be measured.

In terms of measures, the framework should identify:

- The data items necessary to measure these constructs and concepts.
- The populations that should be covered in the measurement of different concepts.

In each of the domains, it would be important to identify the data items that are considered essential in a particular domain and those that are not essential but allow a more complete understanding of the domain. There would also be value in developing model questions to give examples of questions considered to offer good measures of the concepts of interest and to increase comparability between studies.

Finally, any international guidelines regarding the collection of information on literacy in household surveys should also address the range of other sources of error that may affect the overall quality of data related

[31] See also Wagner (1990[65]).

[32] In particular, the implementation of international adult literacy assessments such as IALS, ALL LAMP, PIAAC and STEP.

to literacy collected through household-based studies. In many cases, the major risks to data quality are less measurement error than representation error – e.g. error associated with coverage of the population, sampling and non-response. Beyond the use of valid and reliable measures, the collection of high quality information on literacy also demands adherence to rigorous standards regarding data collection and processing.

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Education (2014), *Standards for Educational and Psychological Testing*, American Educational Research Association, Washington DC. [68]

Bynner, J. and S. Parsons (2006), *New Light on Literacy and Numeracy: Summary Report*, National Research and Development Centre for Adult Literacy and Numeracy, http://www.nrdc.org.uk/?p=315 (accessed on 6 August 2018). [37]

Cellule de Planification et de Statistiques (2014), *Enquête Agricole de Conjoncture Intégrée (ECAI) 2014, Mali, 2014 – 2015*, https://microdata.worldbank.org/index.php/catalog/2583 (accessed on 6 August 2018). [60]

Central Bureau of Statistics (2017), *Living Standards Survey 2010-2011, Third Round, Nepal, 2010 – 2011*, https://microdata.worldbank.org/index.php/catalog/1000 (accessed on 6 August 2018). [56]

Central Bureau of Statistics (BPS) of Indonesia (2012), *Indonesia, National Social Economic Survey of 2012*, https://catalog.ihsn.org/index.php/catalog/3031 (accessed on 6 August 2018). [54]

Centre for Longitudinal Studies (n.d), *1970 British Cohort Study*, https://cls.ucl.ac.uk/cls-studies/1970-british-cohort-study/ (accessed on 27 November 2020). [63]

COSIT and KRSO (2007), *Household Socio-Economic Survey 2006-2007, Iraq, 2006 – 2007*, https://microdata.worldbank.org/index.php/catalog/69 (accessed on 6 August 2018). [50]

Demographic and Health Surveys (DHS) (2020), *Monitoring and Evaluation to Assess and Use Results Demographic and Health Surveys (MEASURE DHS)*, https://microdata.worldbank.org/index.php/catalog/dhs/about. [64]

Demographic and Health Surveys (DHS) (n.d), *DHS Questionnaires*, https://www.dhsprogram.com/Methodology/Survey-Types/DHS-Questionnaires.cfm (accessed on 6 August 2018). [59]

Education Policy and Data Center (2009), "How (well) is Education Measured in Household Surveys? A Comparative Analysis of the Education Modules in 30 Household Surveys from 1996–2005", *IHSN Working Paper*, No. 002, International Household Survey Network, http://www.ihsn.org/sites/default/files/resources/IHSN-WP002.pdf. [22]

Education Policy and Data Center (EPDC) (2009), "Household Survey Guidelines on Education for Use in the Context of the IHSN Question Bank", *EPDC Working Paper*, No. 09- 04, EPDC, https://www.epdc.org/sites/default/files/documents/Household_Survey_Guidelines_on_Education.pdf. [32]

Educational Testing Services (ETS) (2014), *A Guide to Understanding the Literacy Assessment of the STEP Skills Measurement Survey*, Educational Testing Services (ETS). [13]

Ferreira, M. and C. Griffin (1996), *Tanzania Human Resource Development Survey: Final Report Volume I: Main Report*, Population and Human Resources Eastern Africa Department, The World Bank, Washington D.C. [28]

Gaëlle, P. et al. (2014), *STEP Skills Measurement Surveys Innovative Tools for Assessing Skills*, World Bank, Washington D.C., http://microdata.worldbank.org/index.php/catalog/2010/download/32353. [5]

Gal, I. (2016), *Assessment of Adult Numeracy Skills. Background paper prepared for the 2016 Global Education Monitoring Report, Education for People and Planet: Creating Sustainable Futures for All; 2016*, University of Haifa, Israel, https://unesdoc.unesco.org/ark:/48223/pf0000245573. [10]

General Statistics Office (GSO) (2004), *Household Living Standards Survey 2004, Viet Nam*, https://microdata.worldbank.org/index.php/catalog/2370 (accessed on 6 August 2018). [46]

Ghana Statistical Service (GSS) (2004), *Core Welfare Indicator Questionnaire 2003, Ghana, 2003*, https://catalog.ihsn.org/index.php/catalog/60. (accessed on 6 August 2018). [48]

Grotlüschen, A. et al. (2020), "Low literacy in Germany: Results from the second German literacy survey", *European Journal for Research on the Education and Learning of Adults*, Vol. 11/1, pp. 127-143, http://dx.doi.org/10.3384/rela.2000-7426.rela9147. [6]

Groves, R. and L. Lyberg (2010), "Total survey error: Past, present, and future", *Public Opinion Quarterly*, Vol. 74/5, pp. 849–879, https://doi.org/10.1093/poq/nfq065. [41]

ICF (2017), *Demographic and Health Survey Interviewer's Manual*, ICF, Rockville, Maryland, http://www.dhsprogram.com. (accessed on 3 August 2018). [21]

INSD (2017), *Burkina Faso - Enquête multisectorielle continue (2014)*, https://nada.web.ined.fr/index.php/catalog/86. (accessed on 6 August 2018). [49]

Institut National de la Statistique du Cameroun (2016), *Cameroun, Troisième Recensement Général de la Population et de l'Habitat 2005*, http://slmp-550-104.slc.westdc.net/~stat54/nada/index.php/catalog/89 (accessed on 6 August 2018). [52]

Institute of Statistical Social and Economic Research - University of Ghana (2009), *Ghana - Socio-economic Panel Survey*, Institute of Statistical, Social and Economic Research - University of Ghana, http://microdata.worldbank.org/index.php/catalog/2534 (accessed on 6 August 2018). [11]

Kenya National Bureau of Statistics (2007), *Kenya National Adult Literacy Survey Report*, http://statistics.knbs.or.ke/nada/index.php/catalog/58/download/239. [34]

Knight, J. and R. Sabot (1990), *Education, Productivity, and Inequality: The East African Natural Experiment*, World Bank, Washington, D.C, [27]

http://documents.worldbank.org/curated/en/371451468773715604/pdf/multi-page.pdf.

Link, M. (2018), "New data strategies: nonprobability sampling, mobile, big data", *Quality Assurance in Education*, Vol. 26/2, pp. 303-314, http://dx.doi.org/10.1108/QAE-06-2017-0029. [66]

Maddox, B. and L. Esposito (2011), "Sufficiency Re-examined: A Capabilities Perspective on the Assessment of Functional Adult Literacy", *Journal of Development Studies*, Vol. 47/9, pp. 1315-1331, http://dx.doi.org/10.1080/00220388.2010.509788. [20]

MICS (n.d.), *MICS6, Instructions for Interviewers*, UNICEF, https://mics.unicef.org/tools. [12]

Mingat, A., F. Ndem and A. Seurat (2013), "La mesure de l'analphabétisme en question. Le cas de l'Afrique subsaharienne", *Cahiers de la Recherche sur l'Education et les Savoirs*, Vol. 12, pp. 25-47, https://journals.openedition.org/cres/2288. [25]

Moore, J. (1988), "Miscellanea, Self/Proxy Response Status and Survey Response Quality, A Review of the Literature", *Journal of Official Statistics*, Vol. 4/2, pp. 155–172, http://www.jos.nu/Articles/abstract.asp?article=42155. [40]

Mullis, I. and M. Martin (2015), *PIRLS 2016 Assessment Framework*, TIMSS & PIRLS International Study Center, https://timssandpirls.bc.edu/pirls2016/framework.html (accessed on 3 August 2018). [16]

Murray, T., I. Kirsch and L. Jenkins (1998), *Adult Literacy in OECD Countries: Technical Report on the First International Adult Literacy Survey*, US Department of Education, National Centre for Education Statistics, https://nces.ed.gov/pubs98/98053.pdf. [35]

Nath, S. (2007), "Self-Reporting and Test Discrepancy: Evidence from a National Literacy Survey in Bangladesh", *International Review of Education / Internationale Zeitschrift für Erziehungswissenschaft / Revue Internationale de l'Education*, Vol. 53/2, pp. 119-133, http://dx.doi.org/10.2307/27715358. [29]

National Bureau of Statistics Nigeria (2010), *Nigeria, National Literacy Survey, 2010*, http://ghdx.healthdata.org/record/nigeria-national-literacy-survey-2010 (accessed on 6 August 2018). [53]

National Center for Education Statistics (NCES) (n.d), *The International Adult Literacy Survey (IALS)*, https://nces.ed.gov/surveys/ials/ (accessed on 27 November 2020). [62]

National Institute of Statistics Niger (2012), *Niger National Survey on Household Living Conditions and Agriculture 2011, Niger, 2011 – 2012*, https://microdata.worldbank.org/index.php/catalog/2050 (accessed on 6 August 2018). [61]

National Statistical Office Thailand (2010), *Thailand, Population and Housing Census 2010*, http://popcensus.nso.go.th/en/ (accessed on 6 August 2018). [51]

OECD (2019), *Skills Matter: Additional Results from the Survey of Adult Skills*, OECD Skills Studies, OECD Publishing, Paris, https://dx.doi.org/10.1787/1f029d8f-en. [7]

OECD (2016), *Skills Matter: Further Results from the Survey of Adult Skills*, OECD Skills Studies, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264258051-en. [67]

OECD (2016), *The Survey of Adult Skills: Reader's Companion, Second Edition*, OECD Skills Studies, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264258075-en. [17]

OECD (2010), *PISA 2009 Assessment Framework: Key Competencies in Reading, Mathematics and Science*, PISA, OECD Publishing , Paris, https://doi.org/10.1787/9789264062658-en. [18]

OECD/Statistics Canada (2000), *Literacy in the Information Age : Final Report of the International Adult Literacy Survey*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264181762-en. [19]

OECD/Statistics Canada (1995), *Literacy, economy and society: results of the first International Adult ... - Statistics Canada, Organisation for Economic Co-operation and Development*, OECD Publishing, https://books.google.fr/books/about/Literacy_economy_and_society.html?id=_9-eAAAAMAAJ&redir_esc=y (accessed on 3 August 2018). [36]

ORC Macro (2001), *Model "A" Questionnaire with Commentary for High Contraceptive Prevalence Countries: MEASURE DHS + Basic Documentation No. 1*, ORC Macro, Calverton, Maryland, USA, https://www.dhsprogram.com/pubs/pdf/DHSQ4/DHS-IV-Model-A.pdf.pdf. [69]

Programme d'analyse des systèmes éducatifs de la Confemen (PASEC) (2016), *Cadre de Référence des Tests PASEC2014 de Langue et de Mathématiques de Début de Scolarité Primaire*, http://www.pasec.confemen.org/wp-content/uploads/2016/03/PASEC_2014_CADRE_REFERENCE_TEST_2A.pdf. [70]

République de Côte d'Ivoire (2000), *Enquête à Indicateurs Multiples MICS2000*, http://www.ins.ci/EDS%20&%20MICS/MICS2000_Rapport_final.pdf. [39]

RTI International (2015), *Early Grade Reading Assessment (EGRA) Toolkit Second Edition*, United States Agency for International Development, Washington D.C., https://www.globalreadingnetwork.net/resources/early-grade-reading-assessment-egra-toolkit-second-edition. [15]

Schaffner, J. (2005), "Measuring Literacy in Developing Country Household Surveys: Issues and Evidence", Paper commissioned for the EFA Global Monitoring Report 2006, Literacy for Life, http://unesdoc.unesco.org/images/0014/001462/146285e.pdf. [23]

Schaffner, J. (2005), "Subjective and objective measures of literacy: Implications for current results-oriented development initiatives", *International Journal of Educational Development*, Vol. 25, pp. 652-657, http://dx.doi.org/10.1016/j.ijedudev.2005.04.008. [24]

Statistics Sierra Leone (SSL) (2004), *Integrated Household Survey 2003-2004, Sierra Leone, 2003 – 2004*, https://microdata.worldbank.org/index.php/catalog/2942 (accessed on 6 August 2018). [58]

Statistics South Africa - Government of South Africa (2015), *South Africa - General Household Survey 2015*, Statistics South Africa - Government of South Africa, http://microdata.worldbank.org/index.php/catalog/2773 (accessed on 6 August 2018). [55]

Survey Research Centre (2016), *Guidelines for Best Practice in Cross-Cultural Surveys*, Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, https://www.ccsg.isr.umich.edu/images/PDFs/CCSG_Full_Guidelines_2016_Version.pdf. [43]

Tajikistan State Statistical Agency (2009), *Living Standards Survey 2009, Tajikistan, 2009*, https://microdata.worldbank.org/index.php/catalog/73 (accessed on 6 August 2018). [57]

The Uganda Bureau of Statistics (UBOS) (2018), *Uganda National Household Survey 2002/2003*, https://catalog.ihsn.org/catalog/2343/related-materials (accessed on 27 November 2020). [47]

UN Statistics Division (UNstats) (n.d.), *Metadata: Target 4.6.*, UN Statistics Division (UNstats), http://www.uis.unesco.org/Pages/default.aspx (accessed on 3 August 2018). [2]

UNESCO (2008), *Using a Literacy Module in Household Surveys: A Guidebook; 2008*, UNESCO Bangkok, Bangkok, http://unesdoc.unesco.org/images/0016/001619/161938e.pdf. [45]

UNESCO (2007), *Education for All Global Monitoring Report 2008*, UNESCO and Oxford University Press, http://www.ungei.org/resources/files/154743e.pdf. [31]

UNESCO (2006), *Education for All Global Monitoring Report 2007: Strong Foundations, Early Childhood Care and Education*, UNESCO, http://unesdoc.unesco.org/images/0014/001477/147794e.pdf (accessed on 3 August 2018). [30]

UNESCO (2005), *Education for All Global Monitoring Report 2006: Literacy for Life*, UNESCO, http://unesdoc.unesco.org/images/0014/001416/141639e.pdf. [33]

UNESCO Institute for Statistics (UIS) (2019), *SDG 4 Data Digest: How to Produce and Use the Global and Thematic Education Indicators*, UNESCO Institute for Statistics (UIS), http://uis.unesco.org/sites/default/files/documents/sdg4-data-digest-2019-en.pdf. [71]

UNESCO Institute for Statistics (UIS) (2018), *SDG 4 Data Digest 2018: Data to Nurture Learning*, UNESCO Institute for Statistics (UIS), http://uis.unesco.org/sites/default/files/documents/sdg4-data-digest-data-nurture-learning-2018-en.pdf. [72]

UNESCO Institute for Statistics (UIS) (2017), *Implementation in Diverse Settings of the Literacy Assessment and Monitoring Programme (LAMP): Lessons for Sustainable Development Goal 4 (SDG 4)*, UNESCO Institute for Statistics (UIS), http://www.unesco.org/open-access/terms-use-ccbysa-en (accessed on 3 August 2018). [4]

UNESCO Institute for Statistics (UIS) (2008), *International literacy statistics: a review of concepts, methodology and current data; 2008*, UNESCO Institute for Statistics (UIS), http://www.uis.unesco.org (accessed on 3 August 2018). [26]

UNESCO Institute for Statistics (UIS) (n.d.), *General metadata on national literacy*, https://tellmaps.com/uis/literacy/#!/tellmap/-601865091 (accessed on 3 August 2018). [3]

United Nations (2005), *Household Sample Surveys in Developing and Transition Countries*, Department of Economic and Social Affairs Statistics Division, United Nations, New York. [42]

United Nations (UN) (2015), *Principles and Recommendations for Population and Housing Censuses Revision 3*, United Nations (UN), https://unstats.un.org/unsd/demographic/meetings/egm/NewYork/2014/P&R_Revision3.pdf. [14]

United Nations (UN) (2015), *Transforming our world: the 2030 Agenda for Sustainable Development, Resolution adopted by the General Assembly on 25 September 2015, A/RES/70/1*, United Nations (UN), http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E. [1]

United Nations Department of Technical Co-operation for Development and Statistical Office (1989), *Measuring Literacy through Household Surveys: A Technical Study on Literacy Assessment and Related Education Topics through Household Surveys*, United Nations Department of Technical Co-operation for Development and Statistical Office, New York, https://unstats.un.org/unsd/publication/unint/DP_UN_INT_88_X01_10E.pdf. [44]

Wagner, D. (2011), *Smaller, Quicker, Cheaper Improving Learning Assessments for Developing Countries*, UNESCO-IIEP, Paris, https://repository.upenn.edu/cgi/viewcontent.cgi?article=1004&context=literacyorg_chapters. [9]

Wagner, D. (2005), *Monitoring and Measuring Literacy. Background paper for the Education for All Global Monitoring Report 2006: Literacy for Life, 2005*, http://unesdoc.unesco.org/images/0014/001462/146206e.pdf. [8]

Wagner, D. (1990), "Literacy Assessment in the Third World: An Overview and Proposed Schema for Survey Use", *Comparative Education Review*, Vol. 34/1, pp. 112-138, http://dx.doi.org/10.2307/1188558. [65]

World Bank (n.d.), *STEP Skills Measurement Household Survey (Waves 1 and 2)*, https://microdata.worldbank.org/index.php/catalog/step (accessed on 6 August 2018). [38]

Annex A. Examples of questions regarding literacy proficiency

Table A A.1. Examples of questions regarding literacy proficiency: Selected household surveys

Question	Response categories	Respondent	Population	Source
Is [NAME] literate?	Yes, No	Household head	All household members	Viet Nam, Household Living Standards Survey 2004
Can [NAME] read and write in any language?	Neither able to read or write Able to read only Able to read and write	Household head	All household members over 10 years of age with a highest level of education of less than primary	Uganda, The Uganda National Household Survey 2002/03
Can [NAME] read and write in any language?	Yes, No	Household head	All household members aged 15 years or more	Ghana, 2003 Core Welfare Indicators Questionnaire
[NOM] sait-il lire et écrire dans une langue quelconque ? Si oui, laquelle ? a. Français? b. Langue nationale? c. Une autre langue?	Oui, Non	Household head	All household members aged 5 or more	Burkina Faso - Enquête Multisectorielle Continue 2014
Can you read and write?	No I just read I read and write	Household head	All household members aged 6 or more who have never attended school	Iraq, Household Socio-Economic Survey in Iraq (IHSES) 2006 – 2007
Can you read or write Thai? Can you read or write other languages?	Yes, No	Household head	All household members	Thailand, Population and Housing Census 2010
Does person speak, read and write any of these languages? English, French, Arab, Other	Yes, No	Household head	All household members	Cameroun, Troisième Recensement Général de la Population et de l'Habitat 2005
Can you read in English language with understanding? Can you write in English language? Can you read in any other language with understanding? Can you write in any other language?	Yes, No	Household head	All household members aged 5 years or more	Nigeria, National Literacy Survey, 2010

Question	Response categories	Respondent	Population	Source
Can read and write: a. Latin alphabets b. Arabian alphabets c. Other alphabets	Yes, No	Household head	All household members aged 5 years and over	Indonesia, National Social Economic Survey of 2012
"Does [NAME] have difficulty in doing any of the following: a = Writing his/her name b = Reading (e.g. newspapers, magazines, religious books) at least one language c = Filling in a form (e.g. social grant forms) at least one language d = Writing a letter in at least one language e = Calculating/working out how much change he/she should receive when buying something in at least one language f = Reading road signs	1 = No difficulty 2 = Some difficulty 3 = A lot of difficulty 4 = Unable to do 5 = Do not know		All household members aged 5 years and over	South Africa, General Household Survey, 2015
Can [NAME] read a letter? Can [NAME] write a letter?	Yes, No	Household head	All household members aged 5 and over (reading) All household members aged 5 and over who can read (writing)	Nepal Living Standards Survey - ROUND III 2010
Can you read? Can you write?	Yes, easily Yes, with difficulty No	Household head	All household members aged 6 and over	Tajikistan - Living Standards Survey 2009
Can you read a simple letter in English? In what Sierra Leonean language can you read a letter? Can you write a letter in English In what Sierra Leonean language can you write a letter? Can [NAME] do written calculations?	(Yes, No) (None, Mende, Temne, Krio, Other) (Yes, No) (None, Mende, Temne, Krio, Other) (Yes/No)	Household head	All household members aged 5 and over	Sierra Leone, Integrated Household Survey 2003-2004
Can you read the newspaper? Can you write a one page personal letter?	Yes, easily Yes, with difficulty No	Sampled respondent	Respondents with a highest level of education of primary school or below	DHS
Can you read and understand a letter or newspaper easily, with difficulty, or not at all?	Easily With difficulty Not at all	Sampled Respondent	Respondents with a highest level of education of primary school or below	DHS Phase III

Question	Response categories	Respondent	Population	Source
Can [NAME] read a short simple statement? Can [NAME] write a short simple statement?	Yes, easily Yes, with difficulty No	Household head	Household members aged 6 and over with a highest level of education of primary school or less	STEP (waves 1 and 2)
[NOM] peut-il lire un petit texte dans une langue quelconque? [NOM] peut-il écrire une phrase simple dans une langue quelconque?	Oui, Non	Household head	Household members aged 4 and over (reading) Household members aged 4 and over who can read (writing)	Mali - Enquête Agricole de Conjoncture Intégrée 2014
[NOM] peut-il lire un petit texte dans une langue quelconque? [NOM] peut-il écrire une lettre dans une langue quelconque?	Oui, Non	Household head	Household members aged 4 and over (reading) Household members aged 4 and over who can read (writing)	Niger - National Survey on Household Living Conditions and Agriculture 2011
In which languages do you speak, and in which languages do you read and write, well enough to work in a job that requires that language? List of languages	Yes/No	Sampled respondent	Persons aged 15-64 years	STEP (wave 2)
What language can you deal with, and what is your level of proficiency in that language? A - Mother tongue level of speaking, level of reading, level of writing B – Second/third language - level of speaking, level of reading, level of writing	Cannot Weak Medium Good	Household head	All household members aged 6 years or more	Iraq - Household Socio-Economic Survey 2006-2007
How would you rate your current reading skills in LANGUAGE?	Cannot read that language Poor Fair Good Very Good	Sampled Respondent	All respondents	IALS
How would you rate your current writing skills in LANGUAGE?	Cannot write in that language Poor Fair Good Very Good	Sampled Respondent	All respondents	IALS
How would you rate your reading skills in LANGUAGE for your main job? How would you rate your writing skills in LANGUAGE for your main job?	Excellent Good Moderate Poor	Sampled Respondent	Employed respondents	IALS

Question	Response categories	Respondent	Population	Source
How would you rate your mathematical skills for your main job?	No opinion/not applicable			
How would you rate your reading skills in LANGUAGE needed in everyday life? How would you rate your writing skills in LANGUAGE needed in everyday life? How would you rate your mathematical skills needed in everyday life?	Excellent Good Moderate Poor No opinion/not applicable	Sampled Respondent	All respondents	IALS
Can you usually read and understand what is written in a magazine or newspaper? PROBE IF YES: Can you usually read this easily or with difficulty?	Yes- easily Yes – with difficulty No	Sampled Respondent	All respondents	British Cohort Study

Sources:
General Statistics Office (GSO), (2004[46]), Household Living Standards Survey 2004, Vietnam, 2004, https://microdata.worldbank.org/index.php/catalog/2370 (accessed on 6 August 2018).
The Uganda Bureau of Statistics (UBOS) (2018[47]), Uganda National Household Survey 2002/2003 https://catalog.ihsn.org/catalog/2343/related-materials (accessed on 27 November 2020).
Ghana Statistical Service (GSS) (2004[48]), Core Welfare Indicator Questionnaire 2003, Ghana, 2003, https://catalog.ihsn.org/index.php/catalog/60 (accessed on 6 August 2018).
Institut national de la statistique et de la démographie (INSD) (2017[49]), Burkina Faso - Enquête multisectorielle continue (2014), https://nada.web.ined.fr/index.php/catalog/86 (accessed on 6 August 2018).
Central Organization for Statistics and Information Technology (COSIT) and Kurdistan Regional Statistics Office (KRSO) (2007[50]), Household Socio-Economic Survey 2006-2007, Iraq, 2006 – 2007, https://microdata.worldbank.org/index.php/catalog/69 (accessed on 6 August 2018).
National Statistical Office Thailand (2010[51]), Thailand, Population and Housing Census 2010, http://popcensus.nso.go.th/en/ (accessed on 6 August 2018).
Institut National de la Statistique du Cameroun (2016[52]), Cameroun, Troisième Recensement Général de la Population et de l'Habitat 2005, http://slmp-550-104.slc.westdc.net/~stat54/nada/index.php/catalog/89 (accessed on 6 August 2018).
National Bureau of Statistics Nigeria (2010[53]), Nigeria, National Literacy Survey, 2010, http://ghdx.healthdata.org/record/nigeria-national-literacy-survey-2010 (accessed on 6 August 2018).
Central Bureau of Statistics (BPS) of Indonesia (2012[54]), Indonesia, National Social Economic Survey of 2012, https://catalog.ihsn.org/index.php/catalog/3031 (accessed on 6 August 2018).
Statistics South Africa (2015[55]), General Household Survey 2015, South Africa, 2015, https://microdata.worldbank.org/index.php/catalog/2773 (accessed on 6 August 2018).
Central Bureau of Statistics, (2017[56]), Living Standards Survey 2010-2011, Third Round, Nepal, 2010 – 2011, https://microdata.worldbank.org/index.php/catalog/1000 (accessed on 6 August 2018).

Tajikistan State Statistical Agency (2009[57]), Living Standards Survey 2009, Tajikistan, 2009, https://microdata.worldbank.org/index.php/catalog/73 (accessed on 6 August 2018).
Statistics Sierra Leone (SSL) (2004[58]), Integrated Household Survey 2003-2004, Sierra Leone, 2003 – 2004, https://microdata.worldbank.org/index.php/catalog/2942 (accessed on 6 August 2018).
Demographic and Health Surveys (DHS) (n.d[59]): https://www.dhsprogram.com/Methodology/Survey-Types/DHS-Questionnaires.cfm (accessed on 6 August 2018).
World Bank , The STEP Skills Measurement Program, https://microdata.worldbank.org/index.php/catalog/step/about (accessed on 27 November 2020).
Cellule de Planification et de Statistiques (2014[60]). Enquête Agricole de Conjoncture Intégrée (ECAI) 2014, Mali, 2014 – 2015, https://microdata.worldbank.org/index.php/catalog/2583 (accessed on 6 August 2018).
National Institute of Statistics, Niger (2012[61]), Survey and Census Division, National Survey on Household Living Conditions and Agriculture 2011, Niger, 2011 – 2012, https://microdata.worldbank.org/index.php/catalog/2050 (accessed on 6 August 2018).
Central Organization for Statistics and Information Technology (COSIT) and Kurdistan Regional Statistics Office (KRSO) (2007[50]), Household Socio-Economic Survey 2006-2007, Iraq, 2006 – 2007, https://microdata.worldbank.org/index.php/catalog/69 (accessed on 6 August 2018).
National Center for Education Statistics (NCES) (n.d[62]), The International Adult Literacy Survey (IALS), https://nces.ed.gov/surveys/ials/ (accessed on 27 November 2020).
Centre for Longitudinal Studies (n.d[63]), 1970 British Cohort Study, https://cls.ucl.ac.uk/cls-studies/1970-british-cohort-study/ (accessed on 27 November 2020).

Annex B. Data from DHS on respondents who can read a short sentence

Table A B.1. Percentage of respondents who can read a complete short sentence, by number of years of education

Women aged 15-49 years with highest educational attainment at primary school or less

	Number of years of primary education						
	0	1	2	3	4	5	6
	%	%	%	%	%	%	%
Angola 2016	2.23	1.57	4.65	10.13	20.33	31.09	49.87
Benin 2012	0.29	13.36	4.60	6.29	15.95	35.29	48.99
Burkina Faso 2010	1.33	5.28	4.70	5.41	12.15	28.11	50.58
Burundi 2016	18.56	20.91	40.61	67.41	85.77	92.20	95.84
Cameroon 2016	0.63	2.03	4.83	13.73	22.33	40.54	56.49
Chad 2015	0.13	0.40	0.63	1.80	4.06	10.89	17.93
Congo 2015	0.52	0.32	0.68	2.54	10.09	15.05	32.56
Congo Democratic Republic 2014	0.34	1.14	3.63	7.13	13.84	24.48	33.50
Côte d'Ivoire 2012	1.25	3.96	9.99	18.49	34.18	53.67	75.90
Ethiopia 2016	1.34	3.27	6.24	9.13	22.47	39.68	48.06
Gabon 2012	3.13	1.88	8.90	15.29	34.81	51.36	59.43
Gambia 2013	0.29	2.10	0.88	8.47	6.28	10.85	15.47
Ghana 2016	1.62	2.61	3.05	5.30	1.34	11.12	17.03
Guinea 2012	0.03	1.84	2.50	3.38	4.90	12.44	14.17
Kenya 2015	5.45	47.79	21.98	14.43	29.34	48.99	56.85
Lesotho 2014	8.51	10.75	23.98	44.60	59.93	66.35	79.61
Liberia 2016	0.60		3.48	7.32	13.74	22.81	24.88
Malawi 2017	1.16	14.76	5.91	20.98	41.22	55.74	78.04
Mali 2015	0.04	11.65	4.25	2.21	2.40	7.52	29.20
Namibia 2013	11.16	20.25	28.46	37.44	57.09	69.24	74.51
Niger 2012	0.52		2.72	1.73	4.35	23.68	35.04
Nigeria 2015	0.55	3.49	0.75	5.95	9.18	4.28	14.29
Rwanda 2015	4.97	16.55	35.01	58.70	80.41	91.87	94.68
Sierra Leone 2016	0.02	4.48	0.70		2.00	7.53	10.39
Swaziland 2007	13.09	17.86	44.51	53.00	68.20	74.68	79.83
Tanzania 2016	3.34	18.15	19.53	30.61	44.16	50.43	74.04
Togo 2014	0.42	3.02	2.68	3.37	11.43	22.69	32.82
Uganda 2016	2.95	7.84	12.85	16.06	24.78	39.41	54.70
Zambia 2014	1.92	7.63	5.85	6.68	13.93	19.60	36.10
Zimbabwe 2015	10.09	19.79	27.53	35.68	45.79	53.56	64.43

Source: Adapted from (Demographic and Health Surveys (DHS), 2020[64]), The Monitoring and Evaluation to Assess and Use Results Demographic and Health Surveys, https://microdata.worldbank.org/index.php/catalog/dhs (accessed on 27 November 2020).