2006/ED/EFA/MRT/PI/86

United Nations
Educational, Scientific and
Cultural Organization

Background paper prepared for the

Education for All Global Monitoring Report 2006

Literacy for Life

Measuring literacy in developing country household surveys: issues and evidence

Julie Schaffner
2005

Measuring Literacy in Developing Country Household Surveys: Issues and Evidence

Julie Schaffner

The author gratefully acknowledges data preparation assistance from Anthony Keats, and helpful discussions with Aaron Benavot, François Leclercq and Howard White.

The Fletcher School, Tufts University
April 17, 2005

Introduction

In the Education for All initiative, the Millennium Development Goals, and the declaration of the UN Literacy Decade, country governments and international development institutions have expressed commitment to increasing literacy rates among adults and children by 2015. They have furthermore expressed commitment to gathering data for identifying priorities, devising policies and charting progress toward the goals. Doing this well will require accurate methods for measuring literacy rates. Whether seeking to measure literacy at the national or international level, or pursuing research on the determinants or results of literacy, researchers will frequently need to collect good measures of literacy through household surveys.

Most developing country household surveys (and censuses) have elicited information on an individual's literacy by asking either the individual or another respondent within the same household (often the household head) to offer an opinion on whether that individual is or is not literate. This contrasts dramatically with household survey-based measurement of literacy and related basic skills in the OECD countries over the last two decades, which has been characterized by a move away from household reports (which are thought to be quite inaccurate) to direct assessment (through the use of cognitive skills tests) and from dichotomous measures of literacy skills toward measuring literacy in multiple levels or on continuous scales. For example, the International Adult Literacy Survey (IALS) uses a test of performance on up to 114 literacy tasks (only a subset of which is administered to any one individual) to categorize individuals into one of five skill levels in each of three skill domains: prose literacy, document literacy and quantitative literacy (Murray, et al., 1998).

The last decade and a half has seen much discussion of the need for improved literacy measures in developing countries (see, e.g., United Nations Statistical Office, 1989; International Literacy Institute, 2003; Terryn, 2003), albeit probably employing methods somewhat simpler and less costly than those of the OECD's IALS (Wagner, 2003). Some surveys attempting to employ improved methods have taken place, but they have been conducted by diverse research groups operating in a decentralized fashion,

have often been conducted in rushed and resource-constrained ways, and only rarely have given rise to reports providing full details on the process of designing and implementing the assessment tool, costs and results. It is thus not surprising that no consensus has yet been achieved regarding how exactly to improve literacy measurement in development countries.

This report examines recent experience with literacy measurement in developing country household surveys, seeking lessons to guide future survey designers. The first four sections of the report set out the dimensions of choice that survey designers face when creating the tools they will use to elicit literacy information at the individual level, discussing first the choice of literacy concept for which measures are sought and then variants on three broad approaches that may be taken to measuring the desired concepts: direct assessment, indirect assessment by household respondent, and construction of proxy measures employing data on highest grade completed in school. The next two sections set out the criteria that should ideally be weighed when making specific choices about how to incorporate literacy measurement into a household survey, and the way in which survey purposes, empirical evidence and context should influence the weights placed on differing criteria and the resulting survey design choices. The rest of the report then musters as much evidence as possible to enlighten these choices by examining the experiences of the diverse experiments with direct literacy assessment over the last decade and a half. In some cases the report relies on survey reports, while in other cases it presents the results of original analysis of micro data generated by the surveys.

A key theme of the report is that while a shift from household assessments and proxy measurement toward direct assessment of literacy is highly desirable, survey designers should recognize some important limitations in our current understanding of direct assessment. In particular, they must recognize that the quality, ease and comparability of direct literacy assessments falls dramatically as the level and range of literacy skills we hope to measure increase. Direct assessments of whether or not individuals can read a simple sentence aloud (a low level literacy task involving mere "decoding" of written language) can be done simply, probably with reasonable accuracy. By contrast, attempts to assess higher levels of reading skill involving comprehension and interpretation of prose and documents appear often to suffer from serious failings and are inherently difficult to construct in a way that would be comparable across countries or ethno-linguistic groups. It may be possible, with much more concerted and expert effort than has been invested thus far, to develop reliable protocols for construction of tests at these higher literacy levels, but we are still far from knowing how to do this well.

The difficulty, non-comparability and current inaccuracy of direct assessment for most literacy skills beyond simple decoding of text has at least two implications. First, all developing country household survey designers should be encouraged to incorporate simple tests of decoding skills, in which respondents are asked to read aloud a simple sentence written on a "flash card.". This approach is almost as easy as asking for a household response regarding an individual's literacy, and appears to provide a much more accurate answer. (Expert care must, of course, be taken in developing, pre-testing, and translating into multiple languages, the sentences to be employed in the assessments.)

Even though the ideal notion of literacy that survey designers would like to measure may involve higher level skills such as reading comprehension, decoding may be the highest level of reading skills for which they can obtain an accurate measure. Many survey designers will also find it infeasible or undesirable to incorporate more than a very simple test of literacy in their surveys, thus encouraging all survey designers (even those who plan also to test literacy at higher levels) to incorporate a flash card test will facilitate comparisons across groups, places and time periods.

Second, attempts to measure more than simple decoding skills should be viewed as developmental and experimental. They should be avoided unless adequate resources are available for intensive involvement by psychometric experts and extensive pre-testing. Detailed reports on how the measures were developed and implemented, how much the process cost, difficulties encountered in the field, and how the measures performed should be disseminated broadly within the development community. In most cases, survey designers should also keep their objectives modest. Even though ideal concepts of literacy may involve the ability to perform a wide range of "literacy tasks", they should seek to measure well the abilities to perform a small number of tasks rather than seeking to measure many abilities while measuring none of them well. A variety of more specific recommendations are presented at the end of the report.

Literacy Concepts

Literacy Skill Domains. In general, literacy has to do with the possession of skills related to the interpretation or use of written language and symbols, but specific concepts of literacy differ in the specific nature of the skills that must be possessed or used if the individual is to be considered literate. Researchers may want to measure literacy skills in one or more "skill domains" defined by making choices along the following dimensions:

- Whether the required skills relate to reading, writing, oral or written mathematical calculations, or interpretation of visual information other than words.
- Whether the required skills relate to tasks common in the formal schooling context to tasks undertaken frequently in every day life.
- Whether the required skills relate to the use of "any" written language (including many possible mother tongues) or only to the use of a specific language (such as a country's official language of government, or perhaps an international language used for business transactions or for acquiring news and information from neighboring countries).

It is common for survey designers to identify three or four skill domains in which to concentrate their efforts. For example, they may decide to focus on every day skills related to reading in the official language, writing in the official language, reading in any language, and numeracy.

Skill Levels. Within any skill domain, individuals may possess skills in varying degrees. For example, within the skill domain of reading, skills may vary from the most basic, such as being able to identify and sound out individual letters, to decoding of words and sentences, to comprehending sentences and documents, locating information within text, and evaluating the purpose or implications of a text. Survey designers must determine how to map performance with varying tasks into a measure of literacy level within the skill domain.

The most common approach in developing country household surveys has been to treat literacy at the individual level as a dichotomous outcome, with individuals being labeled as either literate or illiterate, depending on whether or not they appear to pass a specified skill threshold in one or more skill domains. For example, individuals might be sorted into groups of "literate" and "illiterate" based on whether or not they can read aloud a simple sentence or whether or not they can read a letter with understanding.

Newer approaches to defining literacy often distinguish more than two significant levels. They may categorize individuals into one of three, four or five skill levels within a skill domain. For example, they may give the label "pre-literate" to the ability to identify letters and sound out words, "basic literacy" to the ability to read a loud a simple sentence and the label "functional literacy" to the ability to read a letter with understanding, while labeling those who cannot identify letters or words as "illiterate." Another alternative is to think of literacy as a continuum, which should be measured with a continuous score within a skill domain.

Complications. Though the definition of literacy is often conceived of as a two-stage problem of first identifying skill domains and then mapping skills to literacy levels within skill domains, in reality it may not be so simple. Especially at higher skill levels, it becomes more difficult to distinguish skill domains. For example, higher order numeracy skills often involve interpretation of word problems, for which reading skills in some language are required. Thus it is difficult to define levels of numeracy skill in a way that is independent of reading skills.

Dimensions of Choice in Direct Assessment

While we would like to think of direct assessment as straightforward observation of literacy skill performance, in fact the design of direct assessments is far from straightforward. Direct assessment of literacy skills involves asking selected respondents to perform specific literacy tasks, recording whether or to what extent they were able to perform each task, and constructing one or more literacy measures based on the full set of test results. Thus survey designers must answer many questions regarding the types of tasks for which to set questions, and the specifics of the questions, the response options, the coding of performance by test administrators and much more.

Within any skill domain, direct assessments may differ along the following broad dimensions:

- How many and which literacy tasks will be tested.

- Whether the tests are intended to sort respondents into a small number of well-defined categories or rate them with a continuous score.
- Whether performance in different skill domains will be rated separately, or whether performance in several domains will be aggregated to construct a single literacy index

For any one literacy task, survey designers must furthermore choose:

- How many test items will be employed per literacy task.
- The specifics of the tasks (e.g. what specific document or piece of prose – employing what sorts of vocabulary and concepts -- will respondents be asked to read, which of the many specific mathematical skills that one might acquire will respondents be asked to perform, in which language will mathematics word problems be rendered).
- In what way will respondents be prompted to respond (e.g. orally or in writing, multiple choice or open-ended).
- What guidance the interviewer is given in recording the outcome (e.g. whether the interviewer merely records the response, or renders a judgment on whether a response was correct or achieved easily)

In the interest of reducing the time and cost associated with literacy measurement, it may be desirable to avoid testing of individuals who are almost surely illiterate or almost surely literate, and to avoid unnecessary testing of individuals at levels that are too high or to low for their likely competence. Thus specific approaches to direct assessment will also be distinguished by:

- What filters or eligibility requirements will be employed to determine which respondents are included in the direct assessments (e.g. they may be applied to all household members over 6 years old, or may be restricted to smaller groups using restrictions such as having completed no more than 6 years of schooling, or having been reported literate by the household head).
- Whether to apply tests to all household members in the filtered groups or randomly select (say) one member in each household for direct testing.
- What protocols to use to determine the level of testing at which the direct assessment should begin for a given respondent, and what protocols should be used to determine whether to proceed to the next higher or next lower level (e.g. some surveys apply the same test to all tested individuals, others sort respondents into initial levels based on highest grade completed in school and then have specific protocols related to test performance for determining whether the individual should proceed to the next higher or next lower testing level, some surveys allow interviewers to exercise discretion in whether to begin testing at the lowest levels or not).

The process of developing and implementing the tools will also play a great role in shaping the ultimate character of the information collected. Important elements of choice here include:

- The mix of experts and users who are involved in test development.
- The nature and extent of pre-testing.
- The extent to which psychometric analysis (on pre-test or pilot data) is employed to identify poorly performing questions, appropriate ordering and ranking of questions, and determine whether the number of test items in each skill domain is adequate to accomplish survey purposes.
- The nature and extent of training given to test administrators and survey administrators.
- The nature of field supervision.
- The coding scheme used for recording details of assessment administration and outcomes.

A final set of choices have to do with how to present the results. A key choice that appears important in several examples below is the following:

- Whether to report literacy levels within individual domains or report only summary literacy measures that aggregate across performance in all domains.

Dimensions of Choice in Assessment by Household Respondents

An alternative to direct assessment that has long been considered more appealing because of its relative simplicity is to measure literacy by asking household respondents to make judgments as to whether or not they or other members within their household can accomplish certain literacy tasks. Perhaps the most common approach is to ask a single household respondent to answer one or more questions of the following sort for each member of the household:

"Can [NAME] read and write a simple sentence in English?"
"Can [NAME] read a newspaper?"
"Can [NAME] sign his own name?"

Even when employing this simple approach to literacy measurement, survey designers face several practical choices:

- How many such questions to employ? (e.g. whether separate questions should be asked regarding reading in the official language and mother tongue, and whether reading, writing and numeracy skills should be assessed in separate questions)
- How to phrase questions so as to identify the class of skills relevant to the literacy judgment? (e.g. what references to make to specific languages or specific tasks)
- For whom to collect these assessments?
- Whether to collect the assessments from the individuals themselves or from a single household respondent?

Dimensions of Choice in Proxy Construction

Historically some countries have constructed national level literacy statistics without attempting to measure literacy directly, whether through direct assessment or the use of household opinions. Instead they have employed survey or census data on highest grade attained, and used to this construct proxy measures of literacy. Individuals having completed at least a specified level of schooling (often 4, 5, 6 or 8 years) are considered literate, while those who have completed less or have never attended school are considered illiterate. The main dimension of choice here is:

- What is the threshold level of highest grade attained at which individuals are considered literate?

Choice Criteria

The previous three sections describe a large matrix of possible approaches to literacy measurement in household surveys. Choices must be made not only between the three broad choices, but also along each of the dimensions that characterize a specific measurement tool within each approach. Ideally, survey designers should weigh the full range of benefits and costs of each of the many possible choices before selecting an approach or set of approaches. This section describes the categories of benefits and costs (or advantages and disadvantages) that should be weighed when comparing measurement approaches, indicating some of the choices in which each criterion is likely to play an especially important role. Good decision-making by survey designers would be greatly facilitated by empirical evidence regarding the sizes of the relative benefits and costs associated with different methods. But empirical evidence on the sizes of the benefits and costs is not, by itself, sufficient for making survey design choices. No single method is likely to be superior in all dimensions, thus survey designers will have to decide how to weigh benefits in some dimensions against costs in others. The next section discusses how survey purpose and context will tend to influence those weights and the resulting survey design choices.

Accuracy in measuring performance of specific literacy tasks. All else equal, researchers would prefer to employ measures that are more clearly valid (i.e. come closer to accurately identifying the relevant levels of the right set of literacy skills) and more reliable (i.e. can be counted on to a greater extent to produce the same measure whenever applied in identical contexts). It is useful to divide the discussion of these accuracy concerns into a discussion of (1) the accuracy with which performance on specific literacy tasks may be measured, and (2) the closeness of correspondence between the literacy tasks that a test assesses and the notions of literacy that survey designers would most like to measure.

When using household responses to measure performance on specific literacy tasks, survey designers must acknowledge the significant potential for inaccuracy arising out of the failure to observe performance directly. Even supposing that survey questions could be phrased sufficiently clearly that survey respondents understood exactly the types

of literacy tasks of interest to the surveyors, they may provide inaccurate answers either because they do not know the answer or because they have an incentive (conscious or unconscious) to misreport performance abilities. Results below suggest that these errors are large and systematic, biasing our understanding of literacy rate levels and of differences in literacy rates across groups.

Even though direct assessments measure performance on specific literacy tasks directly, they may still suffer from inaccuracy, for many reasons. When multiple choice questions are employed, guessing introduces the possibility of false positive, and lack of experience with multiple choice formats introduces the possibility of false negatives. When open-ended questions are employed, the texts and topics chosen may fail to exploit vocabulary and concepts that are common in a local context, and the outside test designer's notion of "correct comprehension" of a story or document may differ from the local notion. When word problems are used to test numeracy skills, difficulties with idiosyncratic vocabulary may introduce false negatives.

The accuracy of direct assessments of specific literacy task skills may be increased in several ways. Using multiple questions to assess performance on a single type of task and then averaging -- or requiring that respondents to obtain 3 out of 4 (say) correct answers to prove ability -- reduces the scope for errors arising out of guessing or out of being misled by specific problematic vocabulary words. Accuracy of direct assessments may also be increased through intensive use of test design expertise and extensive pre-testing and psychometric evaluation.

Inaccuracies in direct assessments seem likely to increase with the level or complexity of the literacy task under consideration. It seems likely, for example, that tests of the ability to decode a simple sentence can be done with reasonable accuracy. Commonalities of life everywhere render it likely that one can identify simple sentences about general life circumstances (such as sentences like "Parents love their children" employed by the Demographic and Health Surveys) that do not present vocabulary problems for functional readers in any context. And, because the sentences involve multiple words, the reading of an entire sentence has a built-in safeguard against guessing and false positives. If a respondent has merely to read most of the sentence to pass the reading test, then the test also has a built in safeguard against false negatives deriving from idiosyncratic vocabulary problems.

As the level or complexity of the literacy task increases, however, inaccuracies are likely to become more important. For example, when assessing reading comprehension it is common to employ multiple choice questions that test comprehension in somewhat artificial ways, requiring respondents to fill in blanks, identify appropriate titles for texts, or identifying particular information within the text. The difficulties introduced by guessing are straightforward: even if all test takers were unable to comprehend, if they may choose from four responses, then we might expect one-quarter of the answers to be correct as a result of random guessing. We will also see evidence below of multiple choice questions yielding what seem to be quite clearly inaccurate test results, as a result of unfamiliarity with the testing mode and difficulties

with local use of vocabulary. When assessing ability to understand and fill in documents, differences from place to place in the nature of the relevant documents may render it difficult to set questions of comparable difficulty that are of comparable relevance to local document literacy requirements.

Closeness of correspondence between targeted literacy tasks and desired notion of literacy. Particular literacy assessment tools may differ not only in the accuracy with which they measure performance in specific literacy tasks, but also in the closeness of correspondence between the literacy tasks that are tested and the notions of literacy the survey designers would most like to measure. Household responses may measure performance on tasks that are different from those intended by survey designers because it is difficult to phrase questions in a sufficiently detailed and nuanced way that household respondents of diverse types and experiences all understand the nature of the tasks on concern to survey designers. To some extent household respondents will define for themselves the types of tasks relevant to the question.

While in principal the designers of direct assessments have the luxury of designing tests that pertain very specifically to the notions of literacy that interest them, they should recognize that they may increase overall accuracy by designing tests that pertain to lower level literacy tasks than they ultimately care about, if such tests can be done with greater accuracy. For example, the notion of literacy with which survey designers are most concerned may involve higher level reading skills related to reading comprehension and interpretation. If, however, their ability to construct accurate measures of performance of such skills is low, they may ultimately end up with a more accurate reflection of the literacy that concerns them by testing performance on a lower level task (decoding of a simple sentence), which can be done with greater accuracy.

Generalizability. All else equal, survey designers would prefer greater confidence that the literacy rates they observe in their sample are a good reflection of literacy rates in some larger context, such as the region or country from which their sample was drawn. Such generalizations can be made with more confidence when the sample to which the tests are administered is indeed a random sample of the relevant population and when the sample size is larger. The higher costs per person associated with some direct assessment methods may, in the face of budget constraints, require smaller and perhaps less representative samples.

Comparability. All else equal, researchers would like to employ literacy assessment methods that allow for greater comparability between their survey results and surveys conducted in other places and times. Growing interest in comparability can be seen in the expansion of household survey data collection programs like the Living Standards Measurement Surveys of the World Bank and the Demographic and Health Surveys funded by USAID.

Comparability would appear much more difficult to achieve in the area of literacy assessment than in the assessment of many other context and outcome quantities of interest in developing country household surveys. Identical questions eliciting household

responses regarding literacy of household members will perform in differing ways from place to place, depending on local needs for literacy skills, local perceptions of surveyors' purposes, and local understanding of what it takes to be literate. Identical direct assessment test questions will perform differently in different places depending on local familiarity with the testing mode, local familiarity with vocabulary and documents, and more. Comparability seems easier to achieve when direct testing the decoding of simple sentences than when testing reading comprehension and other skills.

Costs in time and money. All else equal, survey designers would prefer to execute surveys more cheaply and quickly. Different literacy measurement approaches will tend to differ in at least three cost dimensions: financial cost (per respondent), interview time per household, and time required for developing the measurement instrument. Direct assessments will in general cost more than household responses and construction of proxies based on schooling. Expert opinions, care in translation, and pre-testing are required even to prepare simple flash card tests for using throughout a country. Thus even very simple direct assessments will have greater financial cost and longer survey instrument preparation time than household responses. Any other form of direct assessment will be even costlier in these dimensions, especially it if is done well. The simplest direct assessments need not take much more interview time than household responses, though more advanced direct assessments certainly will. Longer literacy interview and assessment times may mean less time for eliciting information on other subjects, and greater risk of non-response, or greater interviewer fatigue.

Ease and Speed of Use and Interpretation. Even highly accurate measures will be of little use if they are not disseminated to, and easily understood and used by, policy makers and researchers. Thus all else equal, survey designers would prefer to collect information that is easy to use and understand. In general, household responses provide results that are easier to use and (at a superficial level) understand than all but the simplest direct assessments.

The use of the results of direct assessment approaches may be complicated by the following practical choices. Filtering that leads to direct assessment of only a sub-population and protocols that lead to different individuals being administered different tests (in the interest of avoiding unnecessary testing) may require that scores be imputed for tests not taken in the relevant population before calculating literacy rates. We will see an example below in which the method in which the details of the coding sheet on which details of test administration and test results were recorded rendered the data very difficult to understand and use. Direct assessments that produce continuous literacy scores, which may allow surveys to pick up subtle differences in school quality and understanding across regions and over time, may be difficult to use intuitively for describing the current stock of literacy in a region. (For example, it is easier to grasp statements about percentages of the population who can read a simple sentence and comprehend a simple letter than it is to grasp statements about percentages of the population achieving scores in different ranges on a complex 100 point test.)

Weighing choice criteria

Ideally survey designers would like detailed information on the sizes of the advantages and disadvantages of various literacy measurement approaches in each of the dimensions discussed in the previous section. Knowing, for example, whether the accuracy improvement associated with using simple direct assessments rather than household responses is large or small, for example, would shed important light on the choice between those two approaches. Most of the rest of this report is devoted to describing what empirical information may be derived from experience over the last decade and a half with various literacy measurement methods.

It is important to recognize, however, that even with complete empirical information on the relative advantages and disadvantages of all methods, no single best method would emerge. A variety of methods will most likely be found to have appealing attributes as well as disadvantages in other areas. Choices from diverse alternatives can be made only after deciding what weights to place on the various attributes. This section briefly describes how differences in survey purpose and differences in survey context may lead to differences in the weights placed on the choice criteria described above, and thus lead to different survey design choices.

The role of survey purposes. While most household surveys share some structural features in common, they differ greatly in their purposes and coverage. One important distinction is between, on the one hand, surveys whose primary aim is to measure and facilitate the study of literacy, and, on the other hand, surveys for which facilitating measurement and study of literacy is just one of multiple purposes. National Literacy Surveys, such as have been conducted in several countries in the last decade, are examples of the first survey type, while the World Bank's Living Standards Measurement Studies are examples of the latter. When offering recommendations to survey designers it is important to remember the distinction between the two. In multi-purpose surveys the weights placed on limiting the time and money costs of literacy assessment are typically greater than in literacy surveys. As multi-purpose surveys are much more common than single purpose literacy surveys, and much more likely to collect the wide range of information required for subtle research on the behavioral and policy determinants of literacy, it is important to search for literacy assessments that are reasonably accurate, simple and quick, even if they do not allow as rich an examination of the levels of literacy in a population as more elaborate methods might allow.

Another important distinction between surveys is between those whose primary purpose is to deepen understanding of the local context and those seeking to facilitate comparisons across regions and countries. The desire for comparability may lead survey designers to focus on measures for which comparability is most easily achieved. If comparability at some level may be achieved easily, it is reasonable to allow for this even in surveys for which comparability is not a primary aim. Such surveys may certainly go beyond this and also attempt to measure more subtle and less comparable aspects of literacy.

Survey purposes will influence not only the choice between broad approaches to measurement, but also choices regarding how many and which skill domains to consider, and which levels of skill to discriminate between within domains. Surveys primarily concerned with charting the performance of upper primary and secondary school systems may focus on higher levels of literacy skills than those concerned with measuring the stock of basic skills considered important by small-scale urban employers. Surveys concerned with access to social services may focus on document reading and the filling out of forms, while surveys concerned with cultural sustainability may focus on mother tongue prose literacy and surveys concerned with job skill acquisition may focus on the comprehension of technical texts and the identification of specific information within such texts, in the language of commerce. Different approaches may be required to achieve comparable levels of accuracy in the different domains.

The role of context and constraints. Quite obviously, timing and budget are more of an issue in some cases than others. When rapid design and fielding of a survey are required, survey designers should be more wary of employing sophisticated literacy assessment methods that can only be done well after extensive pre-testing and development. When money is a more binding constraint, again survey designers should be reluctant to do more than very simple (and reasonably dependable) literacy measurements. If, as may be the case in many developing country contexts, it is difficult to attract adequate input from experts in literacy assessment tool design, then literacy assessment objectives should be kept modest. (Of course, international institutions interested in obtaining more high quality literacy measures should also endeavor to help a wide range of partners develop this expertise.)

Learning from Household Surveys Employing Multiple Literacy Measures

Ideally, we would like to assemble detailed empirical evidence on the accuracy, comparability, speed and difficulty of development and administration, and ease of use and interpretation for a wide variety of literacy measurement methods. Unfortunately, the empirical evidence is much more limited, shedding light on only a small fraction of the measurement methods that might be developed, and providing only incomplete information on a single survey experiment with each method. Nonetheless, careful study of the experiments that have taken place over the last decade and a half does yield some useful observations and recommendations. In what follows I provide discussions of a variety of survey experiences, first discussing studies for which I have only secondary information (usually a single survey report) and then providing original analysis of surveys for which I was able to obtain the primary data. The main focus of these discussions is on answering the following questions for each case:

- What do we learn about the relative accuracy of household reports, proxy calculations and direct literacy assessment, when it comes to measuring the level of literacy and the size of disparities in literacy across groups?
- What do we learn about the challenges and pitfalls of direct assessment, and the choices that may have important effects on the accuracy and usefulness of the resulting literacy measures?

- What do we learn about the importance of phrasing and the choice of respondent in the design literacy measures based on household responses?
- What do we learn about the schooling threshold at which respondents can be considered literate, and thus about the appropriate construction of schooling-based literacy proxies?

Review of Household Survey-Based Measurement Studies: Zimbabwe and Kenya

Perhaps the earliest developing country household surveys to experiment with new, direct assessment methods for measuring literacy were the Zimbabwe Literacy Survey of 1986, and the Kenya Literacy Survey of 1988, to which reference is made in United States Statistical Office (1989). That document includes as appendices the literacy assessment tools employed in the two surveys.

These surveys included a background questionnaire, self-reports regarding literacy skills and usage, as well as direct assessments. The direct assessments included (in English and a local language) word-picture matching, a simple reading test (in which the respondent reads two simple sentences out loud and the interviewer records the number of words read correctly), a simple writing test (in which the respondent writes two simple sentences and the interviewer records the number of words written correctly), and a test of basic arithmetic skills (in which the respondent solves 10 simple problems involving addition, subtraction and multiplication). I have been able to turn up no details on their implementation or results. (It is unfortunate that detailed reports were not produced and disseminated.)

Allusions to these surveys in later literature suggest that these surveys produced evidence that household- or self-assessments of literacy tend to overstate literacy as measured by direct testing. This is especially interesting, given the relatively low skill level at which literacy skills are directly assessed. It seems likely that at lower levels of literacy skill direct assessment methods are more accurate than household responses. If household responses overstate even low-level literacy, then they are likely to overstate more demanding notions of literacy even more. It is also interesting that survey designers chose to assess literacy at fairly low skill levels. It would be interesting to know what roles feasibility considerations and discussions with employers and policy makers played in deciding to focus on these skill levels.

Review of Household Survey-Based Measurement Studies: Morocco

The earliest household survey containing multiple literacy measures for which a detailed report exists is the Morocco Literacy Survey of 1991-92, which was administered to households in two-thirds of each cluster employed in the Morocco National Survey on Household Living Standards. This survey was a joint project of the Moroccan government and the Living Standards Measurement unit of the World Bank. The methods and the results of the survey are presented in Lavy, et al. (1995). The

literacy assessment module was administered to all household members between the age of 9 and 69, in 2240 households. (Documentation for the survey is available in the World Bank LSMS web pages, but the documents there appear not to be complete. Parts of the literacy assessment tools, and parts of the questionnaire where the literacy assessment results were to be recorded, are missing.) In principle the data are available from the World Bank, but researchers must first obtain permission from the Moroccan government, and I have thus far been unable to obtain permission.

Household response questions included self-judgments on reading and writing skills in French and Arabic. The direct assessment was designed to determine a level of performance in the areas of reading and writing in Arabic and French, document familiarity (making use of envelopes, newspapers, electricity bills, medicine labels and national identify cards), and mental and written math (basic arithmetic, measurement, percents, fractions, decimals and plane geometry).

The development of the direct assessment tool and implementation methods seems to have involved significant investment. The first draft of the tool was built off of experiences of the five-year Morocco Literacy Project (headed by Daniel A. Wagner at the University of Pennsylvania). The tool was improved after pre-testing. The improved tool was then piloted and subjected to additional psychometric testing, before a finalized survey and assessment tool was fielded. To avoid non-response, survey teams worked on a 2 pm to 10 pm schedule each day in the field, making first contact and interviewing any members present in the afternoon, and then returning in the evening to interview those who had been gone to school or work in the afternoon.

The tests were designed to sort respondents into levels of competence in each tested skill domain. In reading (Arabic and French) the levels are defined as follows:

- *No competence demonstrated*. The individual is unable to demonstrate ability to decode or comprehend simple written words.
- *Rudimentary ability*. The individual demonstrates decoding and comprehension of single words, but has difficulty reading sentences or connected test with comprehension.
- *Minimal competence*. The individual demonstrates ability to comprehend simple texts, although with some errors.
- *Complete fundamental competence*. The individual demonstrates ability to comprehend a variety of texts, including a newspaper article, without error.

In writing (Arabic and French), the levels are defined as follows:

- *No competence demonstrated*. The individual is unable to demonstrate ability to write correctly a simple dictated word.
- *Rudimentary minimal competence*. The individual demonstrates ability to write single words without difficulty, and complete sentences but with some errors.
- *Complete fundamental competence*. The individual demonstrates ability to write complete sentences without error.

Comparisons of self-reports and direct tests reveal a strong tendency for self-reports to overstate true literacy, this tendency being greater for individuals with 3 years of schooling or fewer. It is also somewhat greater for males relative to females and for rural residents compared to urban ones, though these differences are more modest. Overall, while 45.5 percent of the sample reported that they were literate (in response to household response questions), only 33.1 percent demonstrated at least basic competence, and only 23.8 percent demonstrated full independent competence. Among individuals with only a few years of primary schooling or less, household assessment literacy rates exceed directly assessed literacy rates by 12 percentage points, reaching even as high as 18 percentage points for rural males with little schooling. Among those who reported themselves to be literate only 72.2 percent demonstrated at least basic competence, and 11.2 percent demonstrated no competence at all. There were very few cases of underestimation of competence by the individuals. According to self-report measures, the male-female difference in literacy rates is about 29.2 percent percentage points by self-report while it is 26.0 by direct assessment. The urban-rural gap in literacy rates is 37 percentage points by self-report and 39.4 by direct assessment.

This case study provides additional evidence of the tendency for household reports to overstate literacy (often greatly). Significant investment seems to have been required to produce direct assessment tools that the researchers deemed satisfactory. While little can be said about detailed problems that may have affected difficulty or accuracy, the approach (which identifies a small number of literacy levels within each domain, rather than a mere dichotomy or a continuous measure) yields easily interpretable and rich results. One wonders, however, about whether much is gained by distinguishing the lowest level of reading literacy (decoding single words but not sentences) from illiteracy. It would be interesting to know whether, for example, employers find it valuable for workers to be able to decode labels (even if they cannot read sentences and text), and whether literacy programs targeting those with rudimentary abilities can be structured differently (and more cheaply) than literacy programs targeting those with no reading literacy skills at all.

Review of Household Survey-Based Measurement Studies: Bangladesh

Perhaps the most interesting attempt at improving household survey literacy measurement in the last decade is a literacy survey for Bangladesh, described in Greaney, et al. (1998). The survey was developed by individuals associated with the World Bank and the Bangladesh Institute of Development Studies, in the early stages of development of methods for evaluating micro credit programs in Bangladesh. The survey employed a rural sample of 1798 households in 29 thanas.

The household responses were in fact self assessments, derived from the questions: “Can you read a letter?” and “Can you write a letter?”

The direct assessment skills tests are included as Appendix 1 to this report. These tests were administered to 5235 individuals at least 11 years old, who had completed less than 10 years of schooling. Those who had completed 10 or more years of schooling were not tested, on the assumption that they had already attained the basic skill levels and would be offended if tested. Those who had completed between 5 and 9 years of schooling were initially tested at the highest level. If they passed, they were assumed able to pass at the lower levels. If they failed, the tests were administered in reverse order until a level of competence was discovered. All others were initially tested at the lowest level. Only if they passed a level at a specified level of competency did they pass on to take the next higher level of testing.

The process by which the direct assessment tool was developed was especially interesting. Researchers brought in a panel of employers, civil servants and educators to help identify the minimal acceptable levels of performance in each of several skills domains. This led to the development of an assessment tool that aimed at relatively simple literacy and numeracy skills (by comparison to other direct assessment examples described below).

The development of the assessment tool began with the quick production of an initial draft, with which the individuals involved could gain some experience with the nature of assessment tool design issues. Personnel involved were then given two weeks of training in standardized test development. This was followed by several rounds of pre-testing and psychometric testing. The report details many lessons learned during pre-testing, which motivated significant changes in the assessment tools, such as dropping of some questions (either because everyone could answer them correctly or because they did not appear to work well), adding others (to increase reliability in the measuring of certain skills), and re-ordering of questions, the distinguishing of written and oral math skills, and the dropping of the oral assessment test (which sought to test individuals' ability to engage in conversation).

Reading skills were rated by use of the following categories:

- *Level 1: Non-reader*. A person who fails to identify at least seven out of eight given individual letters.
- *Level 2: Rudimentary-level reader*. A person who correctly identifies a minimum of seven out of eight given individual letters.
- *Level 3: Beginning reader*. A person who correctly completes six out of eight tasks requiring matching a simple word to a picture, and identifying simple words.
- *Level 4: Minimally competent reader*. A person who correctly answers three out of five literal comprehension questions based on a passage.

[A potential difficulty with translation may be noted in the highest level of this test, as shown in the appendix. The meaning of question 17 on the level 3 reading test is unclear to me.]

Writing skills were rated by use of the following categories:

- *Level 1: Non-writer.* A person who fails to write at least six out of eight named letters.
- *Level 2: Rudimentary-level writer.* A person who correctly writes at least six of eight named letters.
- *Level 3: Beginning writer.* A person who writes at least four out of six dictated simple words and sentences which can be identified (allowing for errors).
- *Level 4: Minimally competent writer.* A person who writes a short (12-word minimum) understandable text (allowing for errors in spelling and grammar) within a 6 minute period when given a pictorial prompt.

In pre-testing it was discovered that many individuals could correctly complete mental arithmetic problems even when they could not do written math. Thus it was decided to test oral and written mathematics separately. Oral math skills were rated by use of the following categories.

- *Level 1: Non-numerate (oral).* A person who fails to correctly answer at least four out of five one-step, simple problems involving the four basic mathematical functions (addition, subtraction, multiplication and division) confined to single-digit and two digit numbers.
- *Level 2: Rudimentary-level numerate (oral).* A person who correctly answers at least four out of five simple problems involving the four basic mathematical functions (addition, subtraction, multiplication and division) confined to single-digit numbers.
- *Level 3: Minimally-competent numerate (oral):* A person who correctly answers at least three out of four routine problems involving the four basic mathematical functions.

Written math skills were rated by use of the following categories:

- *Level 1: Non-numerate (written).* A person who fails to correctly answer at least six out of eight tasks involving recognition of four one-digit and two-digit numbers, writing two one-digit numbers and recognizing two simple geometric shapes.
- *Level 2: Rudimentary-level numerate (written).* A person who correctly answers at least six out of eight tasks involving recognition of four one-digit and two-digit numbers, writing two one-digit numbers, and recognizing two simple geometric shapes.
- *Level 3: Low-level numerate (written).* A person who correctly answers at least three out of four simple computations involving the four basic mathematical functions (addition, subtraction, multiplication, division).
- *Level 4: Minimally-competent numerate (written).* A person who correctly answers at least three out of four problems involving the four basic mathematical functions. [These are word problems.]

It may be noted that one can pass the level 1 test here by knowing how to identify numbers and *either* being able to write numbers *or* being able to identify simple geometric shapes. This kind of implicit equivalence of very disparate skills is a troubling characteristic of most numeracy tests discussed in this report.

As in most other surveys with both household reports and direct assessment, the survey provides evidence of a tendency for household respondents to overstate literacy skills. Of those who indicated that they could read, only 83 percent achieved the minimally competent level in the testing. The report does not provide details on how this bias appears to differ across groups.

This case offers further evidence of the superiority of simple direct assessments over household responses. The description of the assessment tool development process suggests that pre-testing and expert analysis play important roles in shaping the accuracy and usefulness of the resulting data. As in the case of the Morocco study, the aim of distinguishing three or four skill levels within each skill domain appears to allow for two good results: avoidance of unnecessary testing by applying only enough of the assessment tool to determine the highest level at which the respondent can pass (rather than applying the entire test to all individuals), and production of statistics that are richly descriptive while also easily comprehended.

Review of Household Survey-Based Measurement Studies: Botswana

Botwana undertook a literacy assessment survey in 1993. Some description of the survey is provided in Comeyras and Chilisa (2001) and Chilisa (2003), but I have not obtained detailed documentation. It appears that another such survey was undertaken in 2003, but I have been unable to locate any documents describing that survey or its results.

The 1993 survey concerned itself with abilities to read and to write in English and Setswana (but not in other languages), and with numeracy. The direct assessments targeted a fairly low level of skill, which according to Chilisa (2003) is comparable to about that attained in one year of schooling. In reading, respondents had to read aloud two sentences ("The man goes to his farm every day." "He looks after his cattle and crops."), and perform two word-picture matching exercises (choosing one of three words to match a picture of a cup and a donkey). In writing, the test was a simple dictation exercise (sentence not provided in my sources). In numeracy, the respondent had to read numerals aloud, write down numbers, count objects, do written addition and subtraction problems (one each), and read the time indicated on clocks. Individuals were deemed to have the requisite skill if they achieved a 50 percent passing mark on the relevant test. The tests were administered to individuals 12 to 65 years old, who never attended school or left school before completing standard 5.

The tendency for household assessment measures to overstate direct assessment measures was lower here than in most of the other surveys reviewed here, with only 2 percent of those who said they could read or write in English or Setswana failing the

related direct test. They were somewhat more likely to overstate their literacy in Setswana than in English. Success rates were highest in numeracy, following by literacy in Setswana and then English.

This case indicates that the tendency for household reports to greatly overstate literacy (as given by direct assessment) is not universal, as the overstatement is much smaller here. As we will see, this is not merely the result of employing a very low level of literacy testing; even simpler direct assessments produced markedly lower literacy rates than household responses in Ethiopia and Nicaragua. The case also raises the possibility that differences in literacy rates across languages are biased in household reports.

Review of Household Survey-Based Measurement Studies: Ondangwa and Windhoek, Namibia

In an effort to understand better the potential errors associated with using household reports or proxy measures of literacy, and to understand better the literacy rate figures derived from the 1991 Census, Namibia's National Literacy Committee carried out a "small, quick and targeted literacy assessment survey" (Namibia Ministry of Education and University of Namibia, 1994). It was not, strictly speaking, a household survey. It employed non-random samples of individuals in two regions (one urban – Windhoek -- and one rural -- Ondangwa), which included any "learners" who were present at the time of the survey in selected Stage 1 and Stage 3 literacy groups. While the report does not provide a description of the literacy program to which these groups relate, it appears that the groups are led by literacy promoters as part of a national literacy program bringing literacy skills to out of school youth and adults. Stage 1 groups concentrate on basic literacy skills in the mother tongue, while Stage 3 groups concentrate on basic literacy skills in English. The survey designers further limited attention to individuals who had been with the literacy program less than 6 months, because they wanted to examine the relationship between literacy skills and levels of schooling completed prior to entering the literacy program. In total, the sample included just 320 learners, 225 from Stage 1 and 95 from Stage 3. Females constituted a large fraction of the sample. Participants' levels of formal schooling (completed prior to entering the literacy program) ranged from none to some secondary schooling, though most had left school before completing primary school. The two samples differed in both age (the Ondangwa sample being older) and language mix.

For convenience, the survey was administered during four consecutive regular meetings of the literacy groups. On the first day the purpose of the survey was explained. On the second day interviews (including questions eliciting self reports on literacy) were administered. On the third day the learners did practice exercises of the sort that would be used in the direct assessment with help from the literacy promoters. On the fourth day they completed the direct assessments without help.

Self reported literacy figures come from a series of interview questions, regarding whether the respondent can read in any language, read a letter, read a newspaper, read and write in English, write in any language and write a letter.

The direct assessments included three parts, each consisting of a 6-point test. The reading component required respondents to find (in lists) words relevant to a picture, and state where a pictured bus was going (based on a sign in the picture indicating the bus's destination). Respondents were then asked to read a short letter (from a husband writing home to say he has found work and is sending money) and answer three questions about content, which required understanding both language and the structure of letters. The writing component required the respondent to write his or her name, write two words related to a picture, and compose a reply to the simple letter employed in the reading test, using at least three sentences (which had merely to be readable and understandable to obtain full credit). The numeracy component included an oral word problem involving addition and subtraction, two written addition and subtraction problems, two multiplication and division problems, and two questions on time and date in which they identified the time shown on a clock and circled a date on a calendar rendered in English. Respondents were counted as "literate" if they scored 4 or more on *both* reading and writing tests. They were considered "numerate" if they scored 4 or more on the numeracy test.

The study's stated definition of literacy was "the ability to read, in mother tongue, a small paragraph on everyday life, and to write (legibly) at least three sentences communicating an everyday message, in the form of a letter." This does not correspond well to the way in which literacy was defined in practice (scoring 4 or more on both reading and writing tests). One could score 4 or more points on the reading and writing tests just described without being able to demonstrate all the skills described in the official definition of literacy, because 3 out of 4 points on each test related to assessment of lower level skills.

The report concludes that self reports of literacy "grossly overstate" true literacy (as measured by direct testing). By self-report 74 (67) percent of the sample is able to read in any language in Windhoek (Ondangwa). When compared to the percentage of individuals who passed *both* reading and writing tests, which is 58 (35) percent in Windhoek (Ondangwa), the self-reports do indeed appear to grossly overstate literacy.

More detailed test results, presented in Table 1, raise serious questions about the validity of the direct assessments, however. Note first that in both Windhoek and Ondangwa, the percentage of respondents who pass *either* the reading test or the writing test are much higher than the percentage who passed *both* the reading and writing tests. This suggests an unexpectedly poor correlation between scores on reading and writing tests. Note second that while in Windhoek the percentage passing the reading test is higher than the percentage passing the writing test, in Ondangwa the reverse is true. One might speculate that cultural differences (including familiarity with buses and letters) and linguistic differences cause the test to perform differently in different contexts, and

perhaps to label some individuals as illiterate because they lack familiarity with certain conventions or situations, even though they can read and write simple letters.

The percentages of respondents passing the numeracy test were similar in both samples to the percentages passing the reading test.

The report also investigated the errors in using a standard proxy for literacy based on years of schooling (relative to the purportedly correct direct assessment measures). It concludes that the best threshold to use for determining literacy by proxy is whether the individual has more than 4 years of schooling. Even with this best choice of threshold, however, the proxy is not accurate. Among those with 4 years or less, 17 percent are literate by direct assessment, and among those with more than 4 years of schooling, 30 percent fail direct literacy tests. (The report seems to draw an overly positive conclusion from this about the use of the education proxy, stating: "The survey results suggest that the 'education proxy' of a minimum of four years of schooling can be used to sort out those who are probably already literate from those who are not. It does, however, wrongly classify some people as illiterate and also, wrongly, some people as literate. It is possible that these two effects more or less cancel each other out." The study then goes on to use the 4-year threshold for studying literacy rates using the 1991 Census.)

This case provides additional evidence of disparities between household reports and direct assessments. It also provides an example of a quickly developed direct assessment that appears not to have performed well, and an example in which aggregating over performance in two skill domains (in this case, reading and writing) produces statistics that seem to tell quite different stories from the disaggregated statistics.

Review of Household Survey-Based Measurement Studies: Lao PDR

The Lao National Literacy Survey 2001 was administered to 8100 households in 136 of the 142 districts of the Lao People's Democratic Republic (Lao People's Democratic Republic, 2004). A household roster questionnaire administered to the household head elicited basic demographic information for all household members, and household literacy reports for all members at least 6 years old. In-depth interviews were then conducted for one randomly selected person per household fitting the criteria of being between 15 and 59 years of age, not having completed lower secondary school, and (in an unusual variant on screening for the direct assessment) having been reported as literate by the household head. Direct tests in reading, writing, numeracy and "visual literacy" were administered to this sub-sample.

The survey and assessment were drafted by a national technical working group and discussed with UNESCO Bangkok and UNICEF Vientiane. During a one week "training of trainers" two days were spent in the field, pre-testing the questionnaire. Two-week surveyor training sessions were then conducted. The "lessons learned" presented in the report suggest that the process felt rushed and resources constrained to those involved in it.

The household report on literacy, available for all household members at least 6 years old, are the household head's answers to the questions:

- Can he/she read, write and calculate in the Lao language?
- Can he/she read, write and calculate in any other language? (If so, please specify language.)

Notice the aggregation across reading, writing and calculation skills. One wonders what weight respondents would place on the diverse skills in deciding how to respond.

Skill tests in each of the four areas (reading, writing, numeracy and visual literacy) each contain five questions or short sets of questions. The questions are intended to measure skills in three levels. The first two question segments correspond to skills taught in the first two levels of nonformal education or the first three grades in formal school. If the respondent can answer neither of these question segments correctly, the test is terminated. The third question pertains to skills taught in the second to third level of nonformal education and grades 3 or 4. Again, if the respondent cannot get question 3 right, the test is terminated. The last two questions pertain to skills taught in the third level of nonformal education and grades 4 to 5.

The five reading skill question segments involve the following activities:

1. Matching four words to four pictures (hen, cow, mountain, mosquito).
2. Indicating what time a clock tells.
3. Telling who will receive a letter from looking at an addressed envelope. (Interviewers judge whether done correctly and whether done easily or with difficulty.)
4. Figuring out the time and cost of a bus trip from Vientiane to Pakse from a bus schedule.
5. Reading a text (on "different advice for mothers") and identifying the topic and purpose of the text from a list. (Interviewer judges whether correct and whether the respondent read or chose the topic easily or with difficulty.)

Note that one can pass the first level of this test (questions 1 and 2) by being able *either* to match words and pictures *or* telling time from a clock. These skills, which are implicitly being treated as equivalent, are sufficiently diverse that they are usually put in separate skill domains (with time telling being considered a numeracy skill). To pass the second level (question 3), one must both be able to read proper names and know envelope addressing conventions. The final level (questions 4 and 5) again tests quite diverse skills, involving reading comprehension in both bus schedules and prose.

The five writing skill questions involve the following activities:

1. Counting the number of bananas in a picture and writing the number in numerals and words.
2. Writing a short sentence about a favorite fruit.
3. Writing the respondent's own name and current address (village, commune/district and province).
4. Filling in a form requesting admission to grade 1 for a son or nephew.

5. Writing a personal letter expressing the respondent's gratitude to his or her teacher for effort in teaching literacy.

The first level (questions 1 and 2) treats as equivalent counting and writing numbers (usually considered a numeracy skill) and writing a short sentence about a fruit. The other levels relate to diverse skills requiring both ability to write words and understanding of document conventions.

The five numeracy test questions involve the following activities:

1. Answering six problems in addition, subtraction, multiplication and division.
2. Answering a word problem involving addition and subtraction.
3. Answering a word problem involving diagram, addition and conversion between meters and kilometers with decimals.
4. Answering a two-part word problem involving multiplications and notion of average.
5. Answering a word problem involving interest calculation.

The visual literacy skills test involved the following activities:

1. Locating own province on a map.
2. Interpreting a diagram about disease transmission involving pictures and arrows.
3. Drawing arrows between pictures to show a logical sequence.
4. Numbering pictures to show logical sequence.
5. Writing a short story based on a picture.

I find these literacy tasks highly diverse and don't have a clear sense of what this test is supposed to accomplish. Perhaps because the reproductions of the pictures included in the report are not large enough, I also found the pictures and diagrams difficult to interpret.

Each test is scored on a 30 point scale, with six points allocated to each of the five activities (the details for the scoring of each activity are not all given in the report). In each dimension an individual is rated as having "basic literacy skills" if the score is at least 8. The individual has "functional literacy skills" if the score is at least 14, and the individual has "sustained functional literacy skills" if the score is at least 22. The rationale for these thresholds is not stated. It might have made more sense to simply define criteria for passing each level of testing, and then base the respondent's category on the highest level passed, rather than the total score.

The report writers chose to aggregate across skill assessments in the three areas of reading, writing and numeracy to construct an overall "literacy" indicator for comparison to the household reports. That is, an individual is rated as having "basic literacy skills" only if his or her scores in all three skill domains were at least 8, and having functional literacy skills if the score in all three areas were at least 14. Having chosen to aggregate this way, many detailed statistics on skill levels in specific domains are not reported.

The information provided does suggest, however, that the choice to aggregate across the three skill categories of reading, writing and numeracy was unfortunate. While the average reading and writing scores were 23.7 and 19.3, the average numeracy test score was 10.6. This suggests that the overall literacy rate from the aggregated measure is pulled down a great deal by the numeracy scores. Many would consider numeracy and literacy to be independent skills. It seems misleading to report that only 10.6 percent of the population is literate, when a much higher percent can read and write with proficiency.

This case illustrates some of the difficulties of interpretation that arise when direct assessments attempt to test a large number of literacy tasks and then aggregate the resulting information into summary statistics on "literacy". Even within literacy domains, the tests seem to treat as equivalent skills that are very different. More important, the authors chose to aggregate results across skill domains, producing overall "literacy" statistics that seem quite misleading, because of the way in which the aggregated figures are pulled down by low numeracy test scores.

Household Survey Data Analysis: Ethiopia 2000

The Demographic and Health Survey (DHS) program, funded by the U.S. Agency for International Development and administered by ORC Macro, has implemented nearly 200 household surveys in over 70 countries since 1984 (www.measuredhs.com). Most DHS instruments prior to 2000 collected only household reports on literacy. After a significant revision of the model questionnaire in 2000, DHS instruments contain simple direct assessments of reading skills. Respondents are asked to read a simple sentence in his or her mother tongue and the interviewer records whether the respondent was able to read some, all or none of the sentences. Sentences include: "Parents love their children," "Farming is hard work.", "The child is reading a book." and "Children work hard at school." According to DHS documents, the process of revising the questionnaires involved a large number of experts and users from a variety of international organizations. It would be useful to obtain more information on the evidence and experience that was drawn on in changing in the literacy question, but my queries have thus far received no answers.

The Ethiopian DHS of 2000 and the Nicaraguan DHS of 2001 appear to be unique among DHS surveys in containing both household assessments and direct assessments of literacy. This section presents a variety of literacy rate comparisons derived from the Ethiopian DHS of 2000. Many important details regarding the sample and survey instrument are presented in Table 2. For interpreting the results for Ethiopia, and for comparing the results here to those from other surveys, the following points are especially important:

- The direct literacy assessment test is a very simple one, in which the respondent is asked to read a single sentence out loud.

- This direst assessment test relates to a concept of literacy that involves only decoding of written language of the sort relevant to every day life, and not to more advanced skills of reading comprehension, or location of information within documents or prose. Neither does it involve assessment of numeracy skills.
- The household assessment question relates to a slightly higher standard of literacy, which includes ability to both read *and write* a simple sentence.
- The direct literacy assessment test was administered only to adults in a certain age range (15 to 49 for women and 15 to 59 for men) and with 6 or fewer years of schooling.
- While the direct assessment literacy test was administered to all women in the required ranges of age and schooling, they were administered only to men with appropriate characteristics in one out of five households.

Table 3 presents literacy rates – as measured by direct assessment, household assessment and proxy construction (using 4 years of schooling as the marker for literacy), for the entire male and female samples and for various sub-samples. It allows comparison across measurement methods in gross differences in literacy rates across groups.

Table 4 provides the results of multivariate (probit) analysis, relating the probability of an individual being found literate by the direct assessment or household report method to various individual, household and respondent characteristics. If we take the direct assessment to be an accurate measure of true literacy, then the direct assessment probit measures the ceteris paribus effect of the characteristics on true literacy, while the coefficients on the characteristics in the household report probit pick up the ceteris paribus effects of the characteristics on both true literacy *and* the extent of reporting bias.

As is well known, the "units" of raw probit coefficient estimates are impossible to interpret directly. Thus rather than reporting such estimates, I use the estimated raw coefficients to calculate estimated "probability impacts" for a reference individual. Probability impacts are the changes in the predicted probability of literacy implied by the model when the relevant characteristic is changed from zero to one, while holding other characteristics at the level relevant to the "reference individual." A reference individual must be chosen, because the nonlinearity of probit models implies that the predicted probability impacts differ across people with different characteristics. Here I take the reference individual to be someone with no schooling, and with all other characteristics at the sample mean level. The estimated probability impacts represent absolute percentage point differences. (That is, a probability impact of .07 means a 7 percentage point increase in the literacy rate when the characteristic is one rather than zero.)

Tables 3 and 4 examine literacy rate differences across two sets of characteristics. The first set includes gender, years of schooling, location, age, and whether the household contains other educated members. We may be interested in gross differences

across these groups (as seen in Table 3) when setting policy priorities.[1] We will also be interested in estimating the "ceteris paribus" effects of these characteristics. That is, we would like to know what would happen to true literacy rates if we could hold all other characteristics constant, while changing just the one. All else equal, males may be more literate than females because households and education systems discriminate in favor of educating males. All else equal (in terms of individual and household characteristics), urban dwellers may be more literate than rural dwellers, because the supply of education is better in urban areas. On the other hand, it may turn out that literacy rates are higher in urban areas not because of conditions intrinsic to urban areas, but because the urban population is younger and more often male. (That is, literacy rates may be correlated with urban location even though urban location has no ceteris paribus effect.)

The best estimates of ceteris paribus effects on true literacy are found in the direct assessment probit in Table 4. Direct assessments seem likely to be more accurate measures of true literacy than the household reports, and the probit allows us to observe effects of any one characteristic while holding the others constant. The household report probits in Table 4 provide estimates of these effects that may be biased, because the coefficients will pick up the ceteris paribus effects of the characteristics not only on true literacy, but also on the extent of household assessment literacy reporting bias.

It is important to exercise caution in interpreting any evidence of "bias" in the estimation of ceteris paribus effects on true literacy when employing household assessments. Such biases may emerge for two very different reasons. We may discover a greater tendency for households to report males (rather than females) as literate even when they cannot read a simple sentence (and are equally illiterate by direct assessment), either because there is a greater tendency for exaggeration when the individual is male or because males more often have certain pre-literacy skills (such as being able to sound out some letters and words) that give household respondents a reason to believe the individuals are literate even when they cannot read a full sentence.

The biggest message of Table 3 is that household assessments have a strong tendency to overstate literacy as assessed through direct testing, at least in populations with only a few years of schooling or less. This is true even though the direct tests focus on only a low level of reading ability (decoding), and even though the household assessment question focuses on a somewhat higher level of skills than does the direct

[1] One of the characteristics included in this list may require some explanation: whether or not the indiviudal's household contains an educated member (other than the individual). Correlations of literacy among household members are of interest to researchers concerned about the equity of access across households to services available only through exercise of literacy skills. If literates are clustered together within households, then a 50 percent rate of literacy among adult individuals may correspond to only 50 percent of households containing any literacy member. If, however, literates are spread more broadly across households, the same 50 percent individual-level literacy rate may translate into a far larger share of households containing at least one literate member. Our concern in the present study is whether the use of household assessments may tend to bias estimates of the extent to which literacy is concentrated within households. If there is a systematic tendency for greater bias in household assessments by individuals in households with other literate members, for example, then measures of co-variation in literacy among household members would tend to overstate the extent to which literacy is clustered together within households.

assessment (making reference to reading *and writing* rather than to reading only). The absolute differences between household assessed literacy rates and directly assessed literacy rates are largest for the sub-samples pertaining to individuals with 1, 2 and 3 years of schooling completed. Among women with one year of schooling, for example, 59.4 percent are considered literate by household assessments, while only 27.1 pass the simple direct reading test. The absolute difference in rates is even larger for men with one year of schooling, for whom the household assessment rate is 65.5, while the direct assessment rate is only 33.3.

Proxy measures of literacy that treat individuals as literate if they have 4 or more years of schooling and illiterate if they have 3 or fewer years yield literacy rates that are lower than direct assessment rates for all groups (within this population that has 6 or fewer years of formal schooling) except for groups having 4, 5 and 6 years of formal schooling. This reflects that even within the Ethiopian population with 6 or fewer years of schooling, a large share of the sample has only a few years of schooling. All such individuals are counted as illiterate by the proxy measure, though many of them are literate by the direct test.

Consider now the differences in literacy rate disparities across major demographic groups as measured by direct assessment and household report. Gross differences across groups, as reported in Table 3, appear exaggerated in the household reports relative to direct assessments. While the percentage point difference in literacy rates between males and females is 18.7 by direct assessment, it is 26.3 percent by household report. Similarly, the urban-rural difference is 20.8 by direct assessment and 29.1 by household report. Literacy rates differences between younger and older age groups, and between speakers of Amarigna (the language of formal education until a reform in the mid-1990s, when regional languages became the languages of instruction), are also larger in household reports than in direct assessments.

At the same time that the percentage point differences in male-female and urban-rural literacy rate differences are higher by household report than by direct assessment, the proportional male-female and urban-rural differences in literacy rates are smaller in household reports than in direct assessment. That is, the ratio of male to female (urban to rural) literacy rates is 2.49 (3.05) by household report and 2.78 (3.60) by direct assessment. This odd outcome arises as a result of the tremendous difference in average literacy rates produced by the two methods. While a 26.3 percentage point difference between male and female rates is larger in absolute terms than an 18.7 percentage point difference, it represents a smaller share of the female literacy rate, which is 17.7 by household report and 10.5 by direct assessment.

Estimated ceteris paribus effects of being male, having a few years of schooling, living in an urban area and being a speaker of Amarigna also appear to be exaggerated in household reports, as seen in Table 4. Comparing the coefficients in the second column of Table 4 to the first, we find that being male rather than female increases directly assessed literacy rates by 4.6 percentage points, while it increases household assessed literacy rates by 13.1 percentage points. The tendency for household reports to overstate

"true literacy" (as measured by direct assessment) thus appears to be stronger for males than females. We also see that as individuals acquire 1, 2 and 3 years of schooling, household assessment rates rise much more quickly than direct assessment rates, suggesting that individuals come to believe, or at least to be willing to report, that they are literate more rapidly than they actually acquire the ability to read a simple sentence. Though the conditional effect of urban location on directly assessed literacy rates is small and negative, the comparable effect on household assessments is a positive 6.7 percent, suggesting a tendency for urban dwellers in Ethiopia to overstate literacy by more than rural dwellers, even after controlling for their level of schooling. Absolute differences across age groups and speakers of different languages are also more pronounced in household assessments than in direct assessments.

As we might expect, Table 3 demonstrates that individuals in households in which at least one other member has 6 years of schooling (a proxy for literacy level of the rest of the household) are higher than among individuals in households with no other household members. The estimated ceteris paribus effects of having an educated household member on an individual's literacy rate by direct assessment, as seen in Table 4, is virtually zero. This indicates that after controlling for other characteristics (urban location being the main characteristic that is shared within households) this household characteristic has no effect on true literacy. The estimated effect in the household report probit is positive and statistically significant (though not very large), suggesting some tendency for correlations of literacy attainment between household members to be higher in household reports than in direct assessments.

The second set of characteristics included in Table 3 and 4 includes characteristics of the proxies who supplied household assessments of individuals' literacy (e.g. whether the respondent was the individual himself/herself or a third party within the household, and, if a third party, whether the respondent is male or female, and educated or uneducated). We are interested in these characteristics primarily because we would like to know their causal effects on the extent of any biases in household literacy reports as measures of true literacy. If we could identify certain types of respondents that provide more accurate household reports, we might be able to improve the accuracy of household assessments through improved rules for respondent selection. For example, if self reports tend to be more accurate than household assessments provided by third parties within households, then future survey designers might hope to improve accuracy of household literacy assessments by requiring interviewers to address literacy questions to each individual rather than to a single household respondent.

The gross correlations between literacy rates and respondent characteristics of Table 3 shed little light on how to improve the accuracy of household reports. The household report probit regressions of Table 4 come somewhat closer to estimating the ceteris paribus effect of respondent characteristics on household assessment biases, because they at least hold constant a set of observed characteristics that may be correlated with the type of respondent and literacy rates. It remains possible, however, that the estimated respondent characteristic effects in the household assessment probit include not only effects on household reporting biases, but also impacts on true literacy rates arising

out of "endogenous selection" of respondents. (That is, individuals who have unobserved characteristics that tend to increase their probability of true literacy may also tend to have jobs or other responsibilities that render them less or more likely to be selected as household respondent.) If the direct assessments provide accurate measures of true literacy, however, then the coefficients on the respondent characteristics in the direct assessment literacy probit should pick up these endogenous selection effects. [The direct assessments are not provided by these respondents; thus there is no reason other than endogenous selection for these variables to enter the direct assessment probit relationship.] Thus the *difference* between the coefficients on these characteristics in the household assessment probit and in the direct assessment probit should provide estimates of the ceteris paribus effects of the characteristics on reporting bias.

Respondent characteristics pick up very little effect in the direct assessment probit in Table 4. This indicates that there is very little "endogenous selection" inducing a correlation between respondent characteristics and true literacy rates. In light of this result, the coefficients on respondent characteristics in the household assessment probit should pick up only the ceteris paribus effects of the respondent characteristics on the extent of reporting biases. But these effect, too, are very small. Thus there is little evidence that careful choice of respondents could significantly improve the accuracy of household reports.

Finally, consider what we may learn about proxy measures of literacy by examining detailed relationships between literacy (as measured by direct assessment) and highest grade completed in school. Table 5 presents directly assessed literacy rates by highest grade of school completed, within various sub-samples defined by gender, age and location. In all sub-samples, it is only in the group that has completed 6 years of school that the literacy rates exceeds 90 percent. The usual approach to literacy proxy construction is to classify all individuals with at least 4 years of schooling as literate. If the rationale for this is that most individuals should have achieved literacy after 4 years, then this threshold appears to low in the case of Ethiopia. While for all sub-samples shown in Table 5 literacy rates reach 90 percent after the same schooling threshold, the groups do differ in literacy rates at lower levels of schooling, where literacy rates for men exceed those for women, and literacy rates for younger groups exceed those for older groups. Simple proxy measures of literacy thus appear unable to provide accurate pictures of literacy disparities across groups.

This case provides an example in which household reports appear to greatly overstate literacy levels and exaggerate differences in literacy rates across groups. It furthermore provides evidence that the apparent biases are very similar across respondents of different types, so there is little scope to improve the accuracy of household reports by trying to select certain types of household respondents. Finally, it provides an example of great inaccuracy in literacy rate levels and differences across groups when measured by schooling-based proxies. The very simple direct assessment measures literacy at a low level, but seems straightforward, relatively cheap and quick to administer, and likely to measure the low level of literacy with reasonable accuracy.

Household Survey Data Analysis: Nicaragua 2001

The Nicaraguan DHS is very similar in structure, and in the nature of the literacy measures, to the Ethiopian DHS just analyzed. Details are described in Table 6. The main differences to keep in mind are that

- The direct assessment was administered only for women in Nicaragua (while such assessments were available for both men and women in Ethiopia) .
- The typical level of schooling completed is higher in Nicaragua than in Ethiopia. [Even when comparing sub-samples of women with 0 to 6 years of schooling completed, the upper grades within this range are more highly represented in the Nicaraguan case than in the Ethiopian case.]
- The Nicaraguan DHS asks respondents whether they have ever participated in a literacy program, and many of them respond in the affirmative. [I take this into account in the analysis below by distinguishing among those with no formal schooling between those who have and have not attended a literacy course.]

Table 7 presents literacy rate comparisons within various sub-samples for Nicaragua. In general, the differences between household assessments and direct assessments are not as large as in the Ethiopian case, though household assessment based literacy rates are still higher in all groups, and especially high for individuals with only a few years of schooling or less. While there are some interesting patterns in the direct assessment literacy rates themselves, there are few particularly striking patterns of difference in apparent household response biases across groups. Both household and direct assessments indicate that literacy rates are lower among those who have attended literacy programs than among those who have completed first grade, higher in urban areas than in rural areas, higher in younger age cohorts, and higher among Spanish speakers than among others.

In the multivariate context of Table 8, differences in both household and direct assessment based literacy rates across rural and urban areas, and across age groups, largely disappear, indicating that the differences in the previous table were driven largely by differences across these groups in the highest level of schooling completed. Household and direct assessment based literacy probits give quite similar patterns of how literacy rates differ across major groups. As in the Ethiopian case, no clear evidence of differences across respondent types in household response biases emerges.[2]

In Table 8 we see that, as in Ethiopia, the threshold number of years of schooling at which 90 percent of the individuals are literate is the same for all ages and regions. That threshold is, however, much lower in Nicaragua (4 years) relative to Ethiopia (6 years). It would be interesting to know whether the threshold here differs for males and females, but direct literacy assessments for males are lacking.

[2] See the discussion of Table 3 in the section on Ethiopia for a motivation and description of the probit estimation and definition of probability impacts.

This case, especially when compared with the Ethiopia case, demonstrates that the extent of biases associated with using household reports or schooling-based proxies probably differ to a great extent across countries. Though household reports overstate literacy in Nicaragua, they do so by much less than in Ethiopia, and the schooling threshold at which most individuals are literate is much lower here than in Ethiopia. The case provides additional evidence that there is little accuracy to be gained in household reports by choosing particular types of respondents.

Household Survey Data Analysis: Tanzania 1993

The Tanzania Human Resource Survey of 1993 employed a nationally representative sample of 5,000 households in Tanzania. The survey was a joint effort undertaken by the Department of Economics of the University of Dar es Salaam, the Government of Tanzania, and the World Bank (as part of the Living Standards Measurement Survey program), and was funded by the World Bank, the Government of Japan, and the British Overseas Development Agency. Data and documentation are available in the web pages for the World Bank's Living Standards Measurement Study (http://www.worldbank.org/html/prdph/lsms/country/tza/tanzdocs.html). Here are some key distinctive features to keep in mind:

- It includes a wider range of both household assessment-based and direct assessment-based literacy measures.
- The direct reading skill assessment tests reading at a higher skill level than did the very simple DHS reading tests. All reading test questions here require the respondent to read (silently) a short paragraph, and then answer multiple choice comprehension questions, requiring them to locate a particular piece of information in the text (often a particular adjective).
- A test of quantitative skills was also included. Respondents were give a test with 10 multiple choice questions, each question testing a different quantitative skills, ranging from simple addition to the solution of an equation.
- The direct tests are only applied to a relatively small sample and only to children, and the procedure used for determining which children would take the test is not described in the documentation.
- In general, the data also inspire less confidence (than do the DHS datasets), as a result of inconsistencies between how the data collection was supposed, in principle, to have happened and what appears to have happened in practice (as described in Table 10A). (See also Ferreira and Griffin, 1996, p. 20, for concerns about the quality of the field work supervision.)

Details of the survey structure and literacy assessment tools are provided in Tables 10A, B and C.

The direct assessment tests are reported to be shorter versions of what is called "Sabot's test", which was developed for employer surveys conducted in Kenya and Tanzania for the research described in Knight and Sabot (1990). According to that reference, the original test was designed by the Educational Testing Service in Princeton,

New Jersey, based in part on school exit exams at the primary and secondary level in Kenya and Tanzania. Knight and Sabot (1990) indicate that because the cognitive tests were considered experimental and risky, they chose not to concentrate a large share of their resources on it. Thus one gets the impression that the direct assessment tool development process was not fully satisfactory to the authors. While few details are provided in the Tanzania documentation, it appears that a shorter version was extracted from the pre-existing Sabot test, without a great deal of systematic pre-testing and adjustment, despite the passage of time and probable changes in school curriculum. Thus it seems reasonable to take this case as an example of applying a direct assessment tool that has not been well honed. It is also important to recognize that the test focuses on a much higher level of functioning in language and math than do the direct tests in the DHS surveys, or in the Bangladesh survey described above.

Despite the dubious quality of the data, some lessons do emerge from the examination of these data. Consider first Table 11, which examines a variety of literacy rates within sub-groups, within the larger population that includes children 7 to 15 years old who have completed 1 to 7 years of school. Notice that given the nature of the dataset, individuals with no schooling are not included in the analysis. In fact, almost all respondents for whom direct assessments were performed are found to be literate by household assessments, even though there is substantial illiteracy (even by household assessments). This raises the suspicion that in practice interviewers only administered direct assessments to people who showed some sign of being able to read. Within this select population, however, we see that only 72.5 percent were able to get at least 2 out of 6 reading test questions correct, and only 1.7 percent were able to get (at least) 5 of 6 questions correct. (No one got all 6 correct.) I will argue below that the indicator of scoring 2 or more correct is a better measure of literacy in this dataset than the indicator of scoring 5 or more correct. The more disaggregated results, which employ quite small samples, do not manifest any very striking patterns.

What is to be made of what appears to be very poor performance on the direct assessments, even among individuals who report themselves able to read and write? Table 13, which presents the details of how respondents answered each question in the cognitive test, provides some insights into the answer. Focusing for the moment on questions 11 through 16 (the reading test), we see the following. The reading test may be broken into two parts, each set of three questions referring to a different paragraph. Performance on the first part was much worse than performance on the second.

Looking more closely at the first reading passage and questions 11 and 12, the response patterns are suggestive of a poorly designed or pre-tested test. If respondents are simply unable to read and comprehend what they read, then we would expect their guesses to be randomly distributed across the response options. But in fact the great majority of test takers chose identical incorrect responses. In both cases the popular incorrect responses appear earlier in the list of potential responses than the correct responses, and both are responses that could be closely associated in the respondents' minds with the correct response. This suggests that respondents read down the list until they found a response that "fit" with their understanding of what they read, checked it

off, and then went on to the next question. I would strongly suspect that in the culture of the individuals who were tested there is no clear distinction between a meat market and a fish market, and that fully literate individuals could reasonably choose meat market as the correct response to question 11. Similarly, they may assume that fishing boats are row boats, and make the same kind of mistake in question 12, even though they are able to read and comprehend the passage.

Two lessons may be drawn from this apparent imperfection in test design. First, a great deal of careful pre-testing and modifications to the direct assessment tool are probably required in most cases to create tests of reading comprehension that function well within a specific culture. Second, it must be especially difficult to design a single simple reading comprehension test that – with only literal translation into different languages – will function equally well in different cultural contexts.

If one ignores the first passage and focuses only on responses to the second passage and the last three questions, then test performance appears much better. The second passage did not seem to present the same difficulties as the first. This suggests that getting a score of at least 2 out of 6 may be a reasonable marker for literacy with this test.

The direct numeracy assessment that seems most comparable to the direct literacy assessment of 2 or more correct is the indicator of whether respondents got at least 3 of 10 math questions right. While 72.5 percent of respondents got at least 2 of 6 reading test questions correct, only 58.7 percent got at least 3 of 10 math questions correct.

Again, closer examination of detailed responses to the math test questions yields some insights into direct numeracy assessment. Each question requires the exercise of quite different skills, and performance across these skills is highly varied. The skills tested and percent getting the right answer may be summarized as follows:

Question	Skills Tested	Percent Correct
1	Simple arithmetic	67
2	Identification of operation required by word problem	50.5
3	Interpretation of decimals	13.2
4	Calculation of percentages	34.5
5	Knowledge of standard measures and ability to estimate	42.6
6	Translation of word problem into multiplication problem, with subtle role for literacy	21.5
7	Multiplication by a fraction (and interpretation of tricky word problem)	25.9
8	Knowledge of area and perimeter	18.6
9	Understanding of algebraic representations	22.8
10	Solution of an equation	19.8

If basic numeracy requires only the ability to do simple arithmetic operations, then only Question 1 is relevant. Unfortunately, given that is it a multiple choice question with 4 responses, we would expect one-fourth of the answers to be correct even if everyone were guessing, raising great questions about the reliability of a literacy indicator based on this question alone. Our confidence in the accuracy of the measure would be much greater if three comparable questions were asked and we treated anyone who got at least two out of three correct as having the required skills.

The results of Question 6 show the great importance of reading literacy to performance on this math test. The question requires respondents to calculate the total number of kilometers a bus travels in a day, given that is makes 3 round trips between two cities that are 12 kilometers apart. Only 21.5 percent got the answer correct. 50.5 percent, however, provided an answer that multiples the distance by 3 rather than 6, suggesting that they knew how to multiply but did not understand the language of "round trip."

Given the small sample size, I performed a more limited type of multivariate analysis than in the Ethiopian and Nicaraguan cases, the results of which are presented in Table 12. (The rationale for the structure and the definition of the probability impacts reported here is the same as that used for Table 4 in the Ethiopia discussion above.) For each literacy measure, I estimated a probit regression of the indicator on an indicator distinguishing males from females, a continuous years of schooling variable (rather than separate indicators for each year of schooling) and an urban residence indicator. Little is significant in the household assessment measure equations, which is no surprise given that in all groups almost all individuals claim to be literate. Direct assessment literacy (measured with the preferred 2 out of 6 indicator) by contrast, does rise significantly (i.e. by about 7 percentage points) with each year of schooling. Male-female and urban-rural differences are not estimated precisely, though the point estimates suggest potentially important differences. Directly assessed numeracy (by the preferred 3 of 10 indicator) also rises significantly with schooling and urban residence.

This case provides more evidence that household reports tend to overstate literacy rates, though the evidence here is less convincing because the direct assessments inspire much less confidence than in previous cases. Perhaps the most important contribution of this case is the insights it gives into some of the pitfalls of direct assessment of higher levels of literacy skills. Cultural or linguistic nuances seem to have rendered some of the literacy comprehension tests invalid. Use of just single multiple choice questions to measure individual numeracy skills probably renders those assessments unreliable. Unclear protocols or lack of field supervision seems to have led to very inaccurate application of filters regarding which children should be tested.

Household Survey Data Analysis: Ghana 2003

The Ghana Household and School Survey of 2003 was designed to allow comparison to the Ghana Living Standards Measurement Survey of 1988/89, for use in a study evaluating education policies in Ghana. Both the 1988/89 survey and the 2003 survey included direct literacy assessments (as well as local school characteristics). A report describing the 2003 survey and reporting on the comparative results is Operations Evaluations Department (2004). This section analyzes the 2003 data. I have so far been unable to obtain permission to use the 1988/89 data, which is described in Poverty and Human Resources Division (1999).

Details of sample, survey and literacy assessment design are provided in Tables 14A through D. The nature of the tests and the population to which they were administered have several distinguishing characteristics:

- The tests were administered to individuals 9 to 55 years old.
- The tests pertain to three areas: English, local language and math.
- The tests were designed with an eye to producing a continuous literacy score (rather than the creation of a small number of literacy level categories), and with the assumption that regression-based imputation methods would be used to fill in many incomplete scores.
- The testing protocol gave discretion to test administrators to skip low levels of testing, and did not direct them to proceed to lower levels of testing if individuals started with a higher level test and failed.

A three-step structure of testing was applied in each of three areas: English, local language and math. In principle, the three steps are as follows. In the first step respondents were presented with flash cards containing a simple sentence or simple arithmetic problem. Respondents who passed the flash card test were eligible for the simple test in each area. Individuals who passed the simple test in any area with 5 or more correct were administered an advanced test. In practice, interviewers were given the discretion to skip lower levels of testing if they felt someone was obviously literate and would be offended by being asked to complete a flash card or simple test. This discretion was intended by the survey designers to be applied in limited circumstances, largely in the case of school teachers, but the discretion seems to have been exercised more liberally. There was no protocol for checking whether an individual started at a higher level of testing in fact failed, and then testing then at the lower level. The simple and advanced tests were identical in 1988/89 and 2003. The flash card test was added in 2003.

The flash card tests were highly comparable to the simple direct reading test administered in the DHS datasets above. The sentences included “They saw him and left” and “I like to eat but not to drink.” The arithmetic problems included “2+3” and “3+3”. Unfortunately, it turns out that it is impossible to extract from the dataset an accurate indicator of who passed (or were judged by test administrators to be able to pass, though they were not tested) this test and who failed. (I return to this below.)

Individuals who passed the flash card test or were assumed capable of passing the flash card test by test administrators were given an 8-item “simple” test in each area. The

simple reading tests involve reading a short paragraph and answering 8 multiple choice questions about it. Seven of the questions involve straightforward literal comprehension and location of relevant phrases in the text. The last question involves somewhat more synthetic skills, as it asks for the best title to the paragraph. The simple math test involves 8 simple problems in addition, subtraction, multiplication and division, employing numerals (rather than words).

The advanced tests in English and mathematics are included as appendices 2 and 3 to the current report. Like the tests employed in the Tanzania data above, the advanced tests are derived from those employed in the research reported in Knight and Sabot (1990). In fact, the Tanzania questions are a subset of the advanced tests employed here. The advanced tests thus are likely to suffer from some of the same problems described above. The reading test includes 29 multiple choice questions requiring literal comprehension, choice of appropriate words to complete sentences (requiring both comprehension and grammar knowledge), and other questions testing the ability to express the same meaning in different words. The math test covers a wide range of subject areas within mathematics, most of them much more advanced than simple addition, subtraction, multiplication and division. It should be noted that the math test was rendered in English for all test takers. It will thus understate the mathematical skills of some individuals who are numerate but not literate in English.

While the tests themselves (especially the flash card test and the simple test in each area) are very interesting for the current purposes, the way in which the tests were administered, and the way in which the results were coded in the questionnaire, render the dataset much less useful than it might have been. The administration and coding were shaped by the authors' primary interest in constructing continuous test scores based on simple and advanced tests, and their plans to use regression-based imputation methods to fill in incomplete scores on these two tests. They were interested in the flash card test only as a method of preventing unnecessary or offensive testing, and not as a literacy assessment in its own right. The authors were furthermore uninterested in determining the highest level of testing at which each individual could pass. For example, if an individual started testing at the highest level, even if the individual failed that test, the researchers were prepared to impute a score on the lower test, rather than administering the lower test and observing performance directly. A coding scheme that elicited less information about specific tests than might be desired seems to have combined with coding errors by test administrators, to produce a dataset in which it is impossible to reconstruct with accuracy indicators of whether individuals did, or were deemed capable of, passing the flash card test and the simple test.

To explain the difficulties, it is useful to reference Table 14B, which presents the section of the questionnaire in which the testing results were recorded, as well as some numbers describing the distribution of responses to each question. The first four questions (a2a, b, c and d) are intended to be indicators of whether or not the individual took the simple English test, simple local language test, simple math test and Raven's test. The information coded here turns out to be inaccurate and uninteresting. In each case there are at least a small number of observations for which the related test scores are

recorded (in subsequent questions) even though the responses here are negative. More important, these are just indicators of whether the simple test was actually administered or not. Individuals might not have taken the test for several very different reasons that we would like to distinguish: they failed the flash card test, they were assumed so highly literate that they were skipped right to the advanced test, or they were not really tested. The next four questions (A3a and b, a4 and a5) record scores for the four simple tests. An important fact I will use below is that the Raven's test was supposed to be administered to every eligible household member (without any flash card test preceding it). Having a Raven's test score thus seems a reasonable indicator of whether an individual did indeed participate in the testing.

Question A6 codes "reason test not taken." There are several problems with this variable. First, there is just one such question but four tests to which it applies. Individuals may have not taken anywhere from one to four tests, but only one reason for not taking the test is recorded. Second, and more important, the categories are not distinguished well for the current purposes. Response "4" is meant to include both those who failed the flash card test and those who passed the flash card test but found the simple test too hard and stopped. Third, the evidence suggests to me that interviewers did not consistently follow instructions regarding the coding of this variable. Some of the observations that lead me to this conclusion are the following:

- In a large number of cases the response recorded is 10 or "other" when it seems likely the person failed the flash card test (and should have been coded as a 4). I infer this from the fact that a 10 is recorded here in many cases even when some test scores are available, indicating that the person was indeed present and participating.
- There are many more cases coded as 8 or "ineligible" in response to question B5 (which is comparable to A6 but pertains to the advanced tests rather than the simple tests) than to A6. According to the manuals, this code should primarily be used for individuals who are outside the relevant age range. It is used, however, for people who are household members of eligible age. Many who are not coded as ineligible for the simple tests are coded as ineligible for the advance tests, suggesting that some test administrators used this coded for people who failed lower level tests (rather than giving the correct code of 4).
- Some cases are coded as having not taken a test because they were away at school or work (or some other reason indicating they were not present) even though some test scores are non-missing for that person, suggesting that the person was indeed present.

Questions B2 a, b and c are indicators of whether advanced tests were taken in English, local language and mathematics. Again, they are inaccurate and uninteresting. This information can be obtained more accurately by observing whether or not a score is recorded in the relevant one of the subsequent three questions (B3a and b, B4). Finally, B5 is the analog to A6, in which a reason for not taking any or all advanced tests is to be coded. It suffers the same problems described above for A6.

According to Operations Evaluation Department (2004) appendix on analysis of test scores, this information was employed in the following way to produce continuous test scores in each area. Anyone for whom the simple test was recorded as "too hard" (A6=4) was given the score of 0 on the simple English test and a random score between 0 and 4 on the simple math test. (Communication with Howard White suggests that they also treated as having failed the flash card test those for whom the answer to A6 was coded as 8 or 10.) After having done this, a regression of simple test scores on advanced scores (of necessity using only observations for which both are present) was used to impute simple test scores for people who had advanced scores but not simple scores. A regression (allowing for double censoring) of advanced scores on simple scores was used to impute advanced scores for those who had simple scores but not advanced scores. Then the resulting simple and advanced scores were added together for a total score. While such imputed continuous scores are useful for comparison with 1988/89 results, and thus for shedding light on changing school quality over the intervening period, it does not seem to be an approach that satisfies the criterion of producing data that are easy to use and interpret.

Given the comparative purposes of the current report, and also given some of the problems with the data described above, I chose to employ the data in a different way. First, it seems preferable to attempt to construct several intuitive literacy level categories rather than to construct a continuous cognitive test score measure. In principle the test structure lends itself to the follow classification scheme.

> *Level 0:* Cannot read a simple sentence (i.e. failed the flash card test)
> *Level 1:* Can read a simple sentence, but cannot answer 5 of 8 questions testing basic reading comprehension (i.e. passed flash card test but failed the simple reading test with a score of 4 or less)
> *Level 2:* Can read a simple sentence and answer correctly at least 5 of 8 questions testing basic reading comprehension, but cannot answer at least 16 (or some other arbitrary number) of 32 more advanced reading comprehension questions (i.e. passed the simple reading test with a score of five or more, or was deemed able to do so by test administrators, but failed the advanced test).
> *Level 3:* Can read a simple sentence, answer correctly at least 5 of 8 questions testing basic reading comprehension, and answer correctly at least 16 of 32 more advanced reading comprehension questions (i.e. passed the advanced reading test with a score of 16 or more).

The distinctions between levels 0, 1 and 2 are especially interesting. The distinction between passing the simple reading test and passing the more advanced test is less clear. It is possible that people with the same skills could pass one or both. Also, the advanced tests included the items employed in the Tanzania test, and thus contain some of the problems discussed in that case.

In what follows I focus on two levels of literacy in each area (English, local language, and math): passing the flash card test and passing the simple test. I cannot do this perfectly. Given the way the data are coded, some individuals who passed the flash

card test but then found the simple test too difficult to take must be treated as having failed the flash card test. In addition, if test administrators started an individual at the advanced level and that test was failed, there was no protocol (of the sort employed in the Bangladesh example above) to go back and determine whether or not the individual could pass the simple test.

The other difference between the way I employ the data and the way the authors of the 2004 report used the data is in how I determine who was indeed tested, and thus who should be treated as having taken the flash card test and failed. Rather than including as having been tested in a given area only those who either has nonmissing scores for the simple or advanced test or a code of 4 on A6, I include everyone for whom a Raven's test score is present. A better coding scheme could have eliminated much of the ambiguity; and a strict protocol governing when to administer which level of test (and how to proceed to the tests in such a way that the levels of literacy can be determined for each individual) would have eliminated additional ambiguity.

In light of the preceding paragraphs, it should be clear that the results in Tables 15 through 17 should be taken with a large grain of salt. Tables 15A through 15D compare literacy rates by various measures in various sub-groups. Contrary to what is found in most surveys reviewed here, the literacy rates in English derived from the simplest direct assessment test (the flash card test) are *higher* than English literacy rates derived from household assessments. This is true for all primary grade levels of highest grade completed, and more true for younger groups than for older groups. The Operations Evaluation Department (1990) report suggests that the quality of schooling in Ghana was rising. This leads to the conjecture that as the quality of schooling, and the standards to which students are exposed, rise, they become less likely to conclude that they are literate when they have only pre-literacy skills. It is also possible, however, that the discretion given to interviewers to skip the flash card test led to an upward bias in the simple direct literacy measure. If interviewers were liberal in exercising the option to skip the flash card test, then some individuals who would have failed the flash card test would have been given the simple test and would be recorded as having passed the flash card test in my measure.

The household assessment of writing ability in English is noticeably lower than the household assessment of reading ability, suggesting that modest changes in the wording of household literacy assessment questions could have important effects on the literacy rates obtained. No clear pattern of differences in biases between males and females emerges in this dataset.

Local language literacy figures demonstrate patterns more like what has been seen in other surveys, with household assessment based literacy rates much higher than even the simplest direct assessment based literacy rates. The differences widen as the highest level of schooling completed increases, and it seen in all age groups.

The simplest direct measures of math literacy are similar to the household assessment based measures, and higher than the lowest level of directly assessed English

literacy. Household assessment based math literacy rates are lower than direct assessments in younger groups but higher than direct assessments in older groups.

Table 16 reports the results of probit estimation exercises comparable to those presented above for Ethiopia and Nicaragua. In each column a different literacy indicator is treated as the dependent variable. In all columns the regressors are the same. The reported coefficients are probability impacts as described above. Comparison of probability impacts for corresponding household and direct assessment measures provides estimates of the extent to which household response biases are determined by the characteristics. The fact that the coefficients on Male in the first and fifth columns are very similar suggests that there is little reason to believe that the use of household assessment based literacy rates would bias our understanding of the ceteris paribus effect of gender on English literacy. The gender effect is larger for the direct math assessment than for the household assessment, suggesting that the use of household assessment measures would lead to an understatement of the gender effect. Contrary to other cases, the household assessment based literacy rates rise more rapidly with highest grade attained than do the direct assessments. Household assessments appear to overstate urban-rural differences in English and math.

As in the analyses for Ethiopia and Nicaragua, the probits produce no evidence of significant differences in reporting biases between self and third party household assessments, or between individuals in families with and without other educated members.

Table 17 reports rates of having passed the flash card test by highest grade completed, within various subgroups. Unlike the cases of Ethiopia and Nicaragua, in many sub-groups the percentage of individuals who are deemed literate by the simple direct test does not rise to 90 percent even by the end of primary school (when all age cohorts are pooled). They get closer for male than for females and for urban rather than rural dwellers. In younger cohorts, however, the rates do reach 90 percent by the end of primary, consistent with the hypotheses that the quality of schooling has improved and that literacy retention is higher for younger people than older people.

This case produced data from which is was difficult to construct discrete literacy indicators, even though the set of tests employed could have been used for such purposes. Given problems in test administration and questionnaire coding, it seems unlikely that the resulting indicators are accurate, and indeed they produces patterns of comparison between household reports and direct assessments that are very different from those seen in most other datasets. The main lessons from this case lie in highlighting useful steps to take to render data fully informative and useful for construction of intuitive literacy level indicators. It would have been preferable to have a strict protocol for test administration that (a) gave no discretion to test administrators over which levels of tests to administer to specific individuals, (b) directed test administrators to administer lower level tests if respondents started testing at a higher level and failed the test, so that it is possible to describe without imputation the highest level at which every tested individual passed the

testing. The questionnaire form on which the test results were recorded could also have been improved. Whether additional training and supervision efforts, combined with stricter protocols and improved forms could have improved test administrators' accuracy in recording testing information is an open question.

Indonesia Family Life Survey – Wave 3 in 2000

It is worth mentioning briefly that another somewhat different attempt has been made to incorporate cognitive test score information into a household survey. The Indonesia Family Life Survey, a collaborative effort of RAND and the Center for Population and Policy Studies (CPPS) of the University of Gadjah Mada, incorporated test scores by asking individuals whether they had even taken the EBTANAS, a series of standardized tests administered at the end of primary, junior secondary and senior secondary school. If they had taken the test, they were asked to produce a report of their scores. If they could not produce a score sheet, they were asked to recall their test scores. Unfortunately, this approach renders the dataset largely uninteresting for the current purposes. The survey is described in Strauss, et al. (2004). The main problems are:

- As the first EBTANAS is administered in the 6th grade, no cognitive scores are available for individuals with no schooling or less than complete primary schooling.
- As the most recent EBTANAS score was in the past for most individuals, it is not a measure of current skills.
- As for many years the EBTANAS was designed at the regional level, the scores are not necessarily comparable across individuals from different regions.

For these reasons, I did not attempt any more detailed analysis of the data.

Other possible sources of insight about direct testing

The surveys discussed thus far are ones for which I was able to obtain enough information that some insights about literacy measurement concerns seemed possible. They are not a complete list of all attempts at direct literacy measurement in developing countries over the last decade and a half. Web searches and conversations with researchers have turned up references to the following surveys, which appear likely to contain direct literacy assessments:

- Bermuda International Adult Literacy and Life Skills Study 2003 (referenced at http://www.nald.ca/WHATNEW/hnews/2005/statscan2.htm)
- Cambodia National Literacy Survey 1999 (Terryn, 2005).
- Chile International Adult Literacy Survey 1998 (for general information on IALS see http://www.statcan.ca/english/freepub/89-588-XIE/about.htm)
- Jamaica National Literacy Survey 1994 (references at www.undp.org/mdg/Jamaica-April2004.pdf)
- Jamaica Adult Literacy Survey 1999 (see Statistical Institute of Jamaica, 1995)
- Malta National Literacy Survey 1999 (references in working papers listed at http://www.um.edu.mt/pressreleases/2000/literacynews.html)

- Trinidad and Tobago National Literacy Survey 1995 (Bernard and Salim, 1995)
- The Young Lives Project (http://www.younglives.org.uk/)

Researchers interested in pursuing this topic further may wish to pursue these sources of information. Efforts to bring together professionals with field experience in direct literacy assessment may also wish to seek out people involved with these surveys.

Summary observations

On indirect household assessments of literacy

- Indirect assessments by individuals themselves and by third parties within their households (together called "household assessments") have a tendency to overstate "true" literacy as measured by direct assessment, even when direct assessments test only rudimentary reading skills (such as the ability to decode and read aloud a simple sentence). The size of the bias varies from context to context, but in some countries the differences are very large.
- The upward bias of household assessment based literacy rates tends to be greatest among individuals with 1 to 3 years of formal schooling. As a result, the biases are greatest precisely in countries with low current levels of schooling, where EFA efforts are most important.
- These biases may be lower where the quality of schooling is higher, and thus the standards to which people hold themselves when providing household assessments are higher (though more evidence would be required before drawing a strong conclusion on this).
- Where the household assessment biases are large, they tend also to differ in magnitude between men and women (being larger for men), and between rural and urban areas (the sign of the difference varies across places), creating the potential for the use of household assessments to bias our understanding of literacy differentials across groups.
- There is no evidence that self assessments are more or less accurate than assessments by third parties (e.g. primary household respondents) within households.

On direct assessments:

- The skill domains in which literacy assessment designers may wish to consider testing are: reading in official language, writing in official language, reading in local language, writing in local language, oral math and written math. These appear to be distinct competencies.
- Judging from the conclusions of the Bangladesh survey designers' discussions with employers, civil servants and teachers, the crucial levels of competency required within each of these skill domains is at a fairly low level. For example, within reading, the crucial distinction pertains to literal comprehension (rather than to more abstract processing of prose). Within writing the focus is on the ability to perform dictation. Within math the focus is on the elementary

operations of addition, subtraction, multiplication and division. This aims the direct assessments at making skill level distinctions at a lower level than do, for example, direct assessments based on "Sabot's test" (variants of which were used in the Tanzania and Ghana data described above).

- Direct assessments designed to sort respondents into a small number of clearly defined categories of skill level seem more useful than those that aim at giving respondents continuous cognitive skill scores. The DHS surveys for Ethiopia and Tanzania sorted individuals into just two categories in the reading domain: those who can and cannot read aloud a simple sentence. The Bangladesh survey sorted individuals into four categories in the reading domain: those who cannot read at all, those who can identify letters, those who can identify words and match them to pictures, and those who can answer simple literal comprehension questions. The resulting information is easier to interpret and present than the information produced by the Ghana survey, which aimed at giving each respondent a continuous test score. It also does a better job of insuring reliability by using several test questions for each well-defined skill level.
- It appears important for direct assessment designers to establish a clear protocol for test administration without discretion on the part of interview administrators or interviewers. The discretion given to test administrators in the Ghana survey – to skip lower levels of testing when they felt it would be offensive – leads to a variety of ambiguities and uncertainties in the use of the data. The aims of avoiding offense, embarrassment and unnecessary testing seem to be accomplished better by setting up a specific protocol of the sort employed in the Bangladesh survey. Interviewers were instructed to use the respondent's highest grade completed in school to determine the level of testing at which to begin. If the respondent passed the initial level of testing, they proceeded to the next higher level. If the respondent failed the initial level of testing, they proceeded to the next lower level. This continued until the survey clearly identified the highest level at which the respondent was competent.
- Difficulties in using the Ghana survey also point to the great importance of careful, complete and well supervised coding of information regarding which tests were administered and what the results were.
- When working with multiple choice test items, it is important to bear in mind the difficulties of interpreting results when only a single test item is devoted to the measurement of a particular skill. If there are four answers, then even if all respondents guessed randomly, one quarter of them are likely to get the answer right. This makes it very difficult to interpret the results of mathematics tests (like Sabot's math test) in which each problem tests a different skill. It would seem wiser to test a smaller number of skills but devote at least three questions to the measurement of each skill, so that a threshold of 2 out of 3 can be used to identify individuals with the requisite skill.
- Discussions of the process of test development for the Bangladesh and Morocco surveys suggest that extensive pre-testing and analysis by psychometric experts is required. Pre-testing led to large changes in the specifics of the assessments. The discussion of problems with the Tanzania direct assessment tool, which appears

not to have been subjected to extensive pre-testing in the local context, also confirm the importance of extensive pre-testing.

On the construction of proxies based on highest grade completed:

- Sorting individuals into "literate" and "illiterate" categories based on whether or not they have completed a certain number of years of school (whether 4, 5, 6 or some other number) is a highly inaccurate procedure.
- The threshold of highest grade completed at which, say, 90 percent of individuals are literate (by direct assessment) also appears to differ greatly from place to place, rendering it even more difficult to construct good proxies.

Recommendations

- The large discrepancies between household assessments and direct assessments of even low-level reading skills in some places underline the great potential for bias in household assessments and the need to move away from household assessments toward direct assessments in developing country household surveys.
- It is important to recognize, however, the limitations of our current understanding when it comes to directly assessing literacy skills in developing countries. It seems likely that simple skills (such as the ability to read aloud a simple sentence, or the ability to perform simple problems in addition, subtraction, multiplication and division) can be tested reasonably accurately and with relative ease. The testing of higher level skills (such as reading comprehension and the highly diverse range of math skills beyond simple arithmetic) appears to be fraught with many more difficulties.
- Designers of a wide range of survey types should be encouraged to incorporate simple "flash card" tests of reading ability, in which respondents are shown a flash card on which is written a simple sentence about every day life and are asked to read the sentence aloud. Interviewers would be asked to record whether or not the respondent is able to read the sentence (perhaps allowing them to distinguish whether the respondent can read "all" or "part" of the sentence, or whether the respondent reads "easily" or "with difficulty"). Though such tests may not measure the "ideal" level and range of literacy skills for every policy or research purpose, they are not much more complicated to implement than the household assessment literacy questions that are also included in a wide range of surveys, and seem likely to be much more accurate in most cases. Encouraging the inclusion of such assessments would also allow for the collection of comparable literacy figures in a wide variety of places and times. Assessing literacy at higher levels or in a wider range of literacy domains is a much more extensive effort, and much less likely to produce statistics that are truly comparable across places.
- It may also be useful to design and implement simple flash card tests of basic arithmetic skills. A flash card should probably contain three or four simple problems involving addition and subtraction (and perhaps multiplication and

division), to give the test comparable reliability to the reading of (most of) a sentence containing 4 or more words.

- Where literacy assessment is a top priority of a survey, it makes sense to undertake somewhat more involved assessments. The following elements appear important for producing meaningful and useful statistics:
    - Selection of a small number of literacy domains, with separate reporting of literacy levels within domains (rather than reporting of indicators that aggregate across domains).
    - Within each domain, clearly defined levels of competence into which respondents are to be sorted (rather than aiming at a continuous measure that aggregates over a large number of literacy tasks).
    - Concentration of valid and reliable testing of a small number of tasks (often with multiple questions per task type) rather than construction of tests with single questions on each of many diverse tasks. (The temptation to test performance on many diverse tasks, with only a single multiple choice question per task, seems especially strong in numeracy testing. Such tests result in unreliable tests that are difficult to interpret.)
    - Involvement of local employers, civil servants and teachers in determining the important levels and domains for testing. It seems likely that this will lead to concentration on testing at somewhat more rudimentary levels of competence than has been the case in some developing country applications (i.e. those employing Sabot's test).
    - Involvement of experts and use of extensive pre-testing in the development of tests at each level in each skill domain.
    - Clear testing protocol without interviewer discretion, which allows the avoidance of some unnecessary testing and embarrassment, while still fully identifying a highest level of competence for each individual.
    - Clear, complete and well-supervised coding of details regarding the carrying out of testing protocols and the test results.
- A push should be made to make public more information on developing country household surveys employing direct literacy assessments. It would be useful to make public not only the survey and assessment instruments, and the resulting data, but also detailed discussions of the process by which the assessments were developed, experiences with experimental versions employed in pre-testing, and the costs of testing and fielding the assessments.
- It would seem useful to bring together researchers who have first hand experience with direct literacy assessments around the world to share insights about the successes and pitfalls. In efforts to develop protocols for the development of literacy assessments around the world, it is important to devote serious effort not only to theoretical discussions of "what literacy is", but also to practical discussion of "which aspects of literacy can we hope to measure well, and how."

References

Bernard, Godfrey St. and Carol Salim. 1995. "Adult Literacy in Trinidad and Tobago – A Study Based on the Findings of the National Literacy Survey 1995," Institute of Social and Economic Research, University of the West Indies.

Chilisa, B.M. 2003. "Assessment of Low Levels of Literacy in Sub-Saharn Africa: Conceptions and Practices," Paper presented at the conference on "Assessment of Low Levels of Literacy," Lyons, France.

Commeyras, M., and B. Chilisa (2001), 'Assessing Botswana's First National Survey on Literacy with Wagner's Proposed Schema for Surveying Literacy in the "Third World"', *International Journal of Educational Development*, 21, p. 433-446.

Greaney, V., S. R. Khandker and M. Alam (1999), *Bangladesh: Assessing Basic Learning Skills*, Washington, D.C. and Dhaka: World Bank. [Accessible at http://lnweb18.worldbank.org/LO%20Web%20Sites/bangladesh%20Web.nsf/0/6ca267d1e46806434625670c0024fc59?OpenDocument]

International Literacy Institute. 2002. *Towards Guidelines for the Improvement of Literacy Assessment in Developing Countries: Conceptual Dimensions Based on the LAP Project*, University of Pennsylvania.

Kirsch, I.S. and A. Jungeblut. 1986. *Literacy Profiles of America's Young Adults*, final report of the National Assessment for Educational Progress, Princeton, NJ: Educational Testing Service.

Knight, J. and R. Sabot 1990, *Education, Productivity, and Inequality*, Oxford University Press for the World Bank.

Lao People's Democratic Republic. Ministry of Education. Department of Non-formal Education. *Lao National Literacy Survey 2001. Final Report.* Bangkok: UNESCO Asia and Pacific Regional Bureau of Education, 2004.

Lavy, Victor, Jennifer Spratt and Nathalie Leboucher. 1995. "Changing Patterns of Illiteracy in Morocco: Assessment Methods Compared," *Living Standards Measurement Study Working Paper no. 115,* Washington, D.C.: World Bank.

Murray, T. Scott, Irwin S. Kirsch and Lynn B. Jenkins (eds.). 1998. *Adult Literacy in OECD Countries: Technical Report on the First International Adult Literacy Survey*, National Center for Education Statistics.

Namibia Ministry of Education and Culture. Directorate of Adult and Continuing Education and University of Namibia. Social Sciences Division. Multi-Disciplinary research Centre. *Adult Literacy in Ondangwa and Windhoek: A Survey of Adult Learners' Literacy Skills and Education.* August 1994.

Operations Evaluation Department. 2004. "Books, Buildings, and Learning Outcomes: An Impact Evaluation of World Bank Support To Basic Education in Ghana," Report Number 28779, World Bank.

Poverty and Human Resources Division, The World Bank. 1999. "Ghana Living Standards Survey (GLSS) 1987-88 and 1988-89 Basic Information".

Statistical Institute of Jamaica. 1995. National Literacy Survey 1994: Jamaica Final Report Kingston: Jamaica Movement for the Advancement of Literacy.

Strauss, J., K. Beegle, B. Sikoki, A. Dwiyanto, Y. Herawati and F. Witoelar. "The Third Wave of the Indonesia Family Life Survey (IFLS3): Overview and Field Report". March 2004.WR-144/1-NIA/NICHD.

Terryn, Benedicte. 2003. "The Literacy Assessment and Monitoring Programme (LAMP): Background and Approach," UNESCO Institute for Statistics (UIS).

Terryn, Benedicte. 2005. *Small Area Estimation of Literacy Skills in Developing Countries: A Two-Part Model for Cambodia Data.* Master of Science Dissertation. University of Southhampton.

Wagner, Daniel A. 1990. "Literacy Assessment in the Third World: An Overview and Proposed Schema for Survey Use," *Comparative Education Review 34(1),* p.112-138.

Wagner, Daniel A. 2003. "Smaller, Quicker, Cheaper: Alternative Strategies for Literacy Assessment in the U.N. Literacy Decade," *International Journal of Educational Research 39, p.293-309.*

United Nations Statistical Office (1989), *Measuring Literacy through Household Surveys*, Document No. DP/UN/INT-88-X01/10E, New York: United Nations Statistical Office. [Zimbabwe]

Table 1
Literacy Rates by Various Measurement Methods
Survey of Adult Literacy in Ondangwa and Windhoek, Namibia

Method of Measurement	Windhoek Sample	Ondangwa Sample
Education Proxy	40 %	34%
Self-reports		
Can read in any language	74	67
Can read a letter	64	63
Can write in any language	73	58
Can write a letter	65	53
Direct testing		
Passed reading test	76	40
Passed writing test	67	50
Passed both reading and writing tests	58	35
Passed numeracy test	76	40

Source: Namibia Ministry of Education and Culture and University of Namibia, 1994.

Table 2 Ethiopia Demographic and Health Survey 2000 Basic Information	
Household sample design	Stratified random sample representative of entire population except for the nomadic population in two regions. Among areas defined by administrative region and rural/urban status, over-sampled those with smaller populations.
Direct literacy assessment test	The interviewer showed the respondent a simple sentence in the respondent's native language, and said "Now I would like you to read out loud as much of this sentence as you can." According to ORC Macro (2001), the sentences are simple sentences appropriate to the country, such as "Parents love their children," "Farming is hard work.", "The child is reading a book.", or "Children work hard at school." The results are coded as "cannot read at all," "able to read only parts of the sentence," "able to read whole sentence," and "no card with required language." Very few results were coded as "able to read only parts of the sentence." In the analysis below, I count as "literate" only those recorded as able to read the whole sentence. I treat as missing values those for whom no card was available in the required language.
Populations to which direct test administered	Women between 15 and 49 years old who reported having completed 6 or fewer years of schooling; and Men (in every fifth household) between 15 and 59 years old who reported having completed 6 or fewer years of schooling.
Household assessment question	The question "Is (NAME) able to read and write a simple sentence?" was addressed to a primary household respondent.
Population for which household question administered	All household members at least 5 years old.
Samples examined in tables	15,270 women between 15 and 49 years old with 6 or fewer years of schooling; 1762 men between 15 and 59 years old with 6 or fewer years of schooling
Definition of schooling variable	Highest grade completed in formal school
Weights employed	The DHS provides individuals weights with the female and male questionnaires, which include household level adjustments for differences in the probabilities of being included in the sample and individual-level adjustments for non-response within demographic groups. Female weights sum to the number of female observations, while the male weights sum to the number of male observations. I have multiplied the male weights by 5 to account for the fact that males were interviewed in only one fifth of the households.

Table 3
Ethiopia
Demographic and Health Survey 2000
Literacy Rate Comparisons

Main samples	Women 15 to 49 years old with 6 or fewer years of schooling completed				Men 15 to 59 years old with 6 or fewer years of schooling completed			
	Direct assess-ment	House-hold Assess-ment	Four Years of Schooling	No. of Obs.	Direct Assess-ment	House-hold Assess-ment	Four Years of Schooling	No. of Obs.
All	10.5	17.7	6.2	12992	29.2	44.0	15.1	2006
Years of schooling completed:								
0	1.4	4.8	0.0	10695	10.5	21.3	0.0	1325
1	27.1	59.4	0.0	303	33.3	65.5	0.0	89
2	40.5	83.9	0.0	459	42.8	82.4	0.0	134
3	63.5	93.3	0.0	409	70.7	94.7	0.0	112
4	76.3	97.1	100.0	371	84.6	99.9	100.0	118
5	85.8	99.7	100.0	385	87.7	98.5	100.0	107
6	92.9	100.0	100.0	370	93.4	99.8	100.0	121
Location:								
Urban	28.8	43.3	21.6	2477	47.8	66.1	40.6	243
Capital	33.1	54.9	31.0	952	57.1	80.2	51.0	85
Other	27.7	40.4	19.4	1525	45.6	62.8	38.1	158
Rural	8.0	14.2	4.1	10515	27.9	42.4	13.2	1763
Age groups:								
15-24	13.8	21.2	8.4	5129	30.2	46.1	18.4	741
25-34	11.6	20.9	7.1	3820	37.9	51.9	21.5	449
35-49	5.0	9.8	2.4	4043	27.1	42.8	10.4	565
50-59					15.4	26.5	3.5	251
Native Language:								
Amarigna	14.9	23.8	8.1	3486	28.5	43.5	8.8	479
Oromigna	7.1	14.3	5.1	3648	24.4	41.9	15.8	625
Tigrigna	12.3	19.2	7.9	1294	39.0	58.1	14.6	165
Other	9.4	14.6	5.1	4564	35.2	45.1	21.2	737
Household assessment provided by:								

Self	8.8	15.7	5.2	6095	29.6	46.6	13.9	596
Other	11.8	19.1	6.9	6897	29.0	42.8	15.6	1410
Whether household has at least one other member with 6 or more years of education:								
At least one	25.9	39.7	19.5	2834	43.2	65.4	31.8	262
None	7.2	12.9	3.3	10158	27.7	41.7	13.3	1744
Third Party Reporter Characteristics:								
Female no primary	14.3	22.0	8.5	2150	29.3	42.5	14.5	997
Female with primary	33.8	52.6	22.6	335	62.5	82.1	42.6	54
Male no primary	7.9	13.6	4.0	3843	24.6	39.4	14.7	317
Male with primary	28.9	45.0	21.8	569	51.6	69.3	46.0	42

Table 4
Ethiopia DHS
Probability Impact Calculations for
Reference Individual with No Schooling and Other Characteristics Set at Sample Means
Based on Probit Estimates of the Determinants of Direct and Household Assessment-Based Literacy[†]
(Standard Errors in Parentheses)

Dependent Variable	Direct Assessment Literacy Indicator	Household Assessment Literacy Indicator
Predicted Probability of Literacy for Reference Individual	.030	.084
Impacts of the Following Regressors on Predicted Probability for Reference Individual		
Male (relative to female)	0.046*	0.131*
	(0.009)	(0.018)
Years of schooling completed (relative to none)		
1	0.276*	0.552*
	(0.041)	(0.046)
2	0.363*	0.736*
	(0.038)	(0.038)
3	0.621*	0.858*
	(0.040)	(0.020)
4	0.785*	0.908*
	(0.032)	(0.007)
5	0.830*	0.905*
	(0.031)	(0.013)
6	0.892*	0.914*
	(0.024)	(0.007)
Location (relative to rural):		
Capital	-0.017*	0.067*
	(0.005)	(0.026)
Other Urban	-0.005	0.011
	(0.008)	(0.021)
Age groups (relative to 15-24):		
25-34	0.015*	0.049*
	(0.006)	(0.015)

35-49	0.015*	0.037*
	(0.007)	(0.015)
Native Language (relative to Amarigna):		
Oromigna	-0.042*	-0.071*
	(0.006)	(0.012)
Tigrigna	-0.008	-0.012
	(0.005)	(0.013)
Other	-0.028*	-0.084*
	(0.005)	(0.010)
Household assessment provided by (relative to other):		
Self	-0.003	0.016
	(0.007)	(0.015)
Whether household has at least one other member with 6 or more years of education (relative to none)	0.004	0.033*
	(0.007)	(0.017)
Third Party Reporter Characteristics (relative to female with less than 6 years schooling):		
Female with 6+ years school	0.006	0.039
	(0.020)	(0.047)
Male with less than 6 years school	-0.012*	-0.020
	(0.005)	(0.013)
Male with 6+ years school	-0.002	0.010
	(0.010)	(0.034)

* Asterisks following coefficient estimates flag coefficients that are statistically significantly different from zero at the 5 percent confidence level.

† Weights were applied in estimating the probit. The sample includes men and women between the ages 15 and 49 for whom both household and direct assessments of literacy are available. The coefficients presented in the table are not the raw probit coefficients, but probability impacts implied by the estimated model. For example, the probability impact of RHS variable X is defined as $\Phi(\beta_0 + \beta_1 + \overline{R}\beta) - \Phi(\beta_0 + \overline{R}\beta)$, where β_0 is the coefficient on X, $\overline{R}\beta$ is a vector of reference values for all the other right hand side variables multiplied by their coefficients, and $\Phi(.)$ is the standard normal cumulative distribution function. A probability derivative of .27, for example, indicates that (for the reference individual) a one-unit increase in the right hand side variable leads to a 27 percentage point increase in the probability of the individual being found literate by the measure employed as the dependent variable (holding all other right hand side variables constant).

Table 5
Ethiopia DHS
Direct Assessment Literacy Rates by Highest Grade Completed, in Various Subgroups

	Females		Males	
Highest Grade Completed	Direct Literacy Rate	Number of Obs.	Direct Literacy Rate	Number of Obs.
0	1.4	10695	10.3	1101
1	27.1	303	34.1	85
2	40.5	459	43.1	128
3	63.5	409	70.3	109
4	76.3	371	85.5	111
5	85.8	385	87.6	104
6	92.9	370	93.4	117

	Ages 15 to 24		Ages 25 to 34		Ages 35 to 49	
	Direct Literacy Rate	Number of Obs.	Direct Literacy Rate	Number of Obs.	Direct Literacy Rate	Number of Obs.
0	2.0	4218	5.6	3425	6.1	4153
1	32.0	247	21.4	106	48.9	35
2	37.0	329	53.6	174	40.5	84
3	71.1	293	64.4	131	63.2	94
4	81.0	256	83.9	135	83.4	91
5	84.6	257	91.1	145	86.5	87
6	95.3	270	90.0	153	93.9	64

	Rural		Urban	
	Direct Literacy Rate	Number of Obs.	Direct Literacy Rate	Number of Obs.
0	4.4	10327	4.6	1469
1	31.3	324	31.1	64
2	42.9	432	35.4	155
3	68.2	320	64.2	198
4	82.2	276	82.9	206
5	87.4	196	85.7	293
6	93.9	184	91.3	303

Table 6 Nicaragua Demographic and Health Survey 2001 Basic Information	
Household sample design	Stratified random sample representative of entire population. Among areas defined by administrative region and rural/urban status, over-sampled those with smaller populations.
Direct assessment test	The interviewer showed the respondent a simple sentence in the respondent's native language, and said "Now I would like you to read out loud as much of this sentence as you can." According to ORC Macro (2001), the sentences are simple sentences appropriate to the country, such as "Parents love their children," "Farming is hard work.", "The child is reading a book.", or "Children work hard at school." The results are coded as "cannot read at all," "able to read only parts of the sentence," "able to read whole sentence," and "no card with required language." Very few results were coded as "able to read only parts of the sentence." In the analysis below, I count as "literate" only those recorded as able to read the whole sentence. I treat as missing values those for whom no card was available in the required language.
Populations to which direct assessment test administered	Women between 15 and 49 years old who reported having completed 6 or fewer years of schooling.
Household assessment question	"Does (NAME) know how to read and write, read only, or neither read nor write?" Few individuals were reported as able to "read only". I treat as literate by this household assessment measure only those who are reported able to "read and write".
Population for which household assessment question administered	All household members at least 6 years old.
Samples examined in tables	7658 women aged 15 to 49 with 6 or fewer years of schooling
Definition of schooling variable	Highest grade completed in school, supplemented by information on the individual's response to the question: "Have you ever participated in an alphabetization program or any other program that taught how to read and write (not including primary school)?" Among those who report having completed 0 years of school, I distinguish between those who have and have not attended a literacy program.
Weights employed	The DHS provides individuals weights with the female questionnaire, which include household level adjustments for differences in the probabilities of being included in the sample and individual-level adjustments for non-response within demographic groups.

Table 7
Nicaragua
Demographic and Health Survey 2001
Literacy Rate Comparisons

Main samples	Women 15 to 49 years old with 6 or fewer years of schooling completed			
	Direct Assessment	Household Assessment	Proxy (four years of schooling)	No. of Obs.
All	67.0	73.7	46.2	7655
Years of schooling completed:				
0	12.4	16.3	0.0	2262
No lit prog	9.0	12.7	0.0	1746
With Lit prog	28.8	33.2	0.0	400
1	48.3	59.2	0.0	360
2	68.9	89.5	0.0	783
3	87.2	97.4	0.0	916
4	92.4	99.2	100.0	965
5	95.3	99.5	100.0	620
6	97.1	99.7	100.0	1749
Location:				
Urban	77.1	83.4	58.9	2933
Large City	81.7	89.1	66.5	322
Other	75.3	81.1	55.9	2611
Rural	59.0	66.1	36.1	4722
Age groups:				
15-24	75.2	80.8	59.5	2859
25-34	65.6	73.1	43.3	2098
35-49	59.7	66.9	34.9	2698
Native Language:				
Spanish	67.5	74.2	46.4	7502
Other	39.9	46.3	32.5	153
Household assessment provided by:				
Self	64.0	70.4	40.8	3833

Other	69.9	77.0	51.5	3822
Whether household has at least one other member with 6 or more years of education:				
At least one	77.1	84.0	59.5	3785
None	55.2	61.7	30.7	3870
Third Party Reporter Characteristics:				
Female no primary	71.6	78.3	54.4	1473
Female with primary	82.3	89.4	67.7	507
Male no primary	57.1	64.8	34.1	1397
Male with primary	82.3	88.7	67.1	445

Table 8
Nicaragua DHS
Probability Impact Calculations for
Reference Individual with No Schooling and Other Characteristics Set at Sample Means
Based on Probit Estimates of the Determinants of Direct and Household Assessment-Based Literacy[†]

	Direct Assessment Literacy Indicator	Household Assessment Literacy Indicator
Predicted Probability of Literacy for Reference Individual	.093	.139
Impacts of the Following Regressors on Predicted Probability for Reference Individual		
Years of schooling completed (relative to none)		
Literacy program only	0.189*	0.185*
	(0.029)	(0.029)
1	0.397*	0.470*
	(0.031)	(0.034)
2	0.600*	0.766*
	(0.023)	(0.016)
3	0.784*	0.836*
	(0.016)	(0.013)
4	0.834*	0.854*
	(0.014)	(0.012)
5	0.861*	0.858*
	(0.015)	(0.012)
6	0.878*	0.858*
	(0.011)	(0.012)
Location (relative to rural):		
Large City	0.016	0.096*
	(0.023)	(0.047)
Other Urban	0.015	0.013
	(0.011)	(0.017)
Age groups (relative to 15-24):		
25-34	-0.004	0.018
	(0.011)	(0.019)
35-49	0.003	0.017
	(0.010)	(0.016)

Native Language (relative to Spanish):		
Other	-0.066*	-0.117*
	(0.012)	(0.017)
Household assessment provided by (relative to other):		
Self	0.015	0.016
	(0.012)	(0.022)
Whether household has at least one other member with 6 or more years of education (relative to none)	-0.001	0.027
	(0.009)	(0.015)
Third Party Reporter Characteristics (relative to female with less than 6 years schooling):		
Female with 6+ years school	0.016	0.078
	(0.028)	(0.074)
Male with less than 6 years school	0.006	0.029
	(0.014)	(0.026)
Male with 6+ years school	0.031	0.064
	(0.025)	(0.049)

* Asterisks following coefficient estimates flag coefficients that are statistically significantly different from zero at the 5 percent confidence level.

† See note to Table 4 for definition of probability impact.

Table 9
Nicaragua DHS
Direct Assessment Literacy Rates by Highest Grade Completed, in Various Subgroups

	Ages 15 to 24		Ages 25 to 34		Ages 35 to 49	
	Direct Literacy Rate	Number of Obs.	Direct Literacy Rate	Number of Obs.	Direct Literacy Rate	Number of Obs.
0	9.0	544	12.5	588	15.1	1014
No Lit.	8.7	503	7.4	490	10.2	753
Lit.Prog.	13.0	41	36.8	98	28.2	261
1	42.4	106	43.2	104	57.7	129
2	73.0	229	71.7	231	64.1	287
3	86.8	310	86.6	260	88.9	311
4	92.0	377	91.3	256	94.6	285
5	94.2	323	95.5	145	98.0	131
6	97.3	841	97.0	427	96.9	413

	Rural		Urban	
	Direct Literacy Rate	Number of Obs.	Direct Literacy Rate	Number of Obs.
0	10.8	1604	17.9	542
No. Lit.	7.8	1335	12.3	411
Lit. Prog.	25.4	269	34.5	131
1	46.3	238	53.8	101
2	68.5	516	69.8	231
3	88.1	560	86.6	321
4	92.1	545	93.4	373
5	94.2	329	96.5	270
6	97.3	720	97.0	961

Table 10A Tanzania Human Resource Development Survey 1993 Basic Information	
Household sample design	Stratified random sample representative of entire population (except for two clusters that were inaccessible by flooding).
Direct assessment tests	The direct literacy assessment involved a 6-question reading test, which was adapted from a longer test developed by the Educational Testing Service in Princeton, New Jersey, and described in an appendix to Knight and Sabot (1990). The six questions, which are aimed at measuring reading comprehension and location of relevant information within prose, are shown in Table 9b. The questions are multiple choice with four possible responses. By random guesses one might expect to get one quarter of the answers right (or at least 1 right out of six). No one answered all six questions correctly. In the analysis here I use two direct literacy assessment measures: whether the individual answered at least two questions correctly, and whether the individual answered at least 5 questions correctly. Missing answers are treated as incorrect responses. 11 individuals included in the test data file provided no answers to any of the literacy questions. They are treated as having provided all incorrect responses. Direct numeracy assessments are derived from a 10-question math test, the details of which are presented in table 9c below. The questions are multiple choice with four possible responses. By random guessing one might expect to get 2 responses correct. The last two questions involve algebra, which seems a very high standard of math literacy for primary schoolers. In the analysis below I use two direct assessment numeracy measures: whether the individual answered at least 3 questions correctly, and whether the individual answered at least 6 questions correctly.
Populations to which direct assessment was administered	According to the survey documentation, the test was administered to a sub-sample of children between the ages of 7 and 15 that were present in the home at the time of the interview. As the test was administered only to 500 children in this age range, while the sample includes roughly 7100 children in this age range, the test was presumably intended to be administered to a random sample of children in this group. In practice, there appear to be some discrepancies between the principle and practice of determining the eligible population for this test. First, though the test was to be applied only to children between 7 and 15 years old, 5 individuals in the test data file have recorded ages younger than this and 294 were older (some very much older). This may be the result of incorrect id codes and matching with household roster information, but seems more likely to be the result of administration of the test to individuals outside the designated age range (as

	a result of very casual assessment of eligibility for the test). Second, though the test was to be applied to all children in this age range, regardless of whether they had attended school, it appears that in practice the test was administered almost entirely to children who had been to school. Though only 67 percent of children ages 7 to 15 mentioned in the household roster had ever been to school, 97 percent of the children included in the test score data file had been to school. Finally, the test was much more likely to be administered to children who were 12 to 15 years old than to children who were 7 to 11 years old. (Testing rates rise quite steadily with age.) Thus it appears that the sample to which the test was administered was not a random sub-sample of children. In the analysis that follows, attention is restricted to children between the ages of 7 and 15 who have ever attended school, and to whom the test was administered. Observations were also dropped if years of schooling information was missing, or if years of schooling was greater than 7 years (as there were only a few such observations, and some of them seemed infeasible for 15 year olds).
Household assessment questions	Answers to the three questions: "Can [name] read a newspaper?" "Can [name] write a letter?" "Can [name] do written calculations?" are used to create household assessment based reading, writing and calculation indicators.
Population for which household assessment questions administered	All household members.
Samples examined in tables	A sub-sample (probably not random) of 479 children between the ages of 7 and 15 who report having completed between 1 and 7 years of schooling.
Definition of schooling variable	Highest grade completed in formal school
Weights employed	A weight derived from the inverse probability of the household's inclusion in the sample.

Table 10B
Reading Test Administered in Tanzania Human Resource Development Survey 1993

CLUSTER NUMBER:		HOUSEHOLD NUMBER:		RESPONDENT ID CODE:	

FOR QUESTIONS 11-16 READ THE PASSAGE BELOW.

You could smell the fish market long before you could see it. As you come closer you could hear merchants calling out about fresh catches or housewives arguing about prices. Soon you could see the market itself, brightly lit and colorful. You could see fishing boats coming in, their decks covered with silver-grey fish.

11	**What kind of market is described above?**		
	(a) . . . . . A vegetable market (b) . . . . . A meat market (c) . . . . . A fish market (d) . . . . . A fruit market	ANSWER	
12	**What could you see coming in?**		
	(a) . . . . . Tug boats (b) . . . . . Row boats (c) . . . . . Fishing boats (d) . . . . . Sailboats	ANSWER	
13	**What covered the decks of the boats?**		
	(a) . . . . . Rope (b) . . . . . People (c) . . . . . Boxes (d) . . . . . Fish	ANSWER	

The cat brushed against the old man. He did not move. He only stood, staring up at the window of the house. The party inside looked warm and friendly, but no one noticed him. The old man walked sadly on, followed by the cat.

14	**What kind of animal was following the old man?**		
	(a) . . . . . Mouse (b) . . . . . Dog (c) . . . . . Cat (d) . . . . . Bird	ANSWER	
15	**What was inside the house?**		
	(a) . . . . . A party (b) . . . . . Some dogs (c) . . . . . An old lady (d) . . . . . A meeting	ANSWER	
16	**The man is described as being**		
	(a) . . . . . Old (b) . . . . . Young (c) . . . . . Thin (d) . . . . . Small	ANSWER	

TIME FINISHED

Table 10C
Math Test in Tanzania Human Resource Development Survey 1993

CLUSTER NUMBER:		HOUSEHOLD NUMBER:		RESPONDENT ID CODE:	

TIME STARTED

1	105 - 16 = ?		
	(a) . . . . . 89 (b) . . . . . 91 (c) . . . . . 99 (d) 111	ANSWER	
2	There are 4 rows of chairs, and 12 chairs in each row. What is the total number of chairs?		
	(a) . . . . . 12 + 4 (b) . . . . . 12 - 4 (c) . . . . . 12 x 4 (d) . . . . . 12 - 4	ANSWER	
3	Which of the following numbers is the largest?		
	(a) . . . . . 0.1 (b) . . . . . 0.01 (c) . . . . . 0.111 (d) 0.1101	ANSWER	
4	1% of 400 is:		
	(a) . . . . . 1 (b) . . . . . 4 (c) . . . . . 40 (d) 400	ANSWER	
5	The height of a man is closest to 2 _____		
	(a) . . . . . millimeters (b) . . . . . meters (c) . . . . . kilometers (d) . . . . . centimeters	ANSWER	
6	Two cities are 12 kilometers apart. Each day, a bus makes 3 round trips between these cities. How many Kilometers does the bus travel each day?		
	(a) . . . . . 72 (b) . . . . . 36 (c) . . . . . 15 (d) 4	ANSWER	
7	An island has an area of about 300 square miles. The government reports that the western 1/3 of the island is not suitable for cultivation. How many square miles of this island are suitable for cultivation?		
	(a) . . . . . 50 (b) . . . . . 100 (c) . . . . . 150 (d) 200	ANSWER	
8	In an office building, each office has about 22 square meters of floor space. In this building, a square office would measure about how many meters on each side?		
	(a) . . . . . 4.7 (b) . . . . . 5.5 (c) . . . . . 11 (d) 484	ANSWER	
9	If a=-3 and b=3, then $2a+b^2$=		
	(a) . . . . . 0 (b) . . . . . 3 (c) . . . . . 9 (d) 12	ANSWER	
10	If 2x - 3 = 17, then x = ?		
	(a) . . . . . 7 (b) . . . . . 10 (c) . . . . . 14 (d) 20	ANSWER	

Table 11
Tanzania
Human Resource Development Survey 1993
Literacy Rate Comparisons
Children 7 to 15 Years Old Who Have Completed 1 to 7 Years of School

	Household Assessment Reading	Houshold Assessment Writing	Household Assessment Calculation	Direct Reading – at least 2 of 6 right	Direct Reading – at least 5 of 6 right	Direct Calculation – at least 3 of 10 right	Direct Calculation – at least 6 of 10 right	Proxy Literacy (Schooling >4)	Number of Observations
All	98.3	96.9	96.5	72.5	1.7	58.7	8.5	66.3	479
Girls	99.0	97.8	98.5	75.8	1.6	56.3	7.7	66.9	273
Boys	97.5	95.9	94.3	68.5	1.7	61.6	9.4	65.7	206
Highest Grade Completed:									
1	81.2	72.3	76.6	74.5	0.0	63.8	1.5	0.0	18
2	98.8	95.1	88.0	69.9	0.0	52.3	11.4	0.0	51
3	97.9	97.3	96.6	49.4	2.1	43.6	2.0	0.0	91
4	100.0	98.4	99.9	65.0	3.5	54.9	9.9	100.0	102
5	100.0	98.7	99.2	81.1	0.1	60.0	7.7	100.0	99
6	100.0	100.0	100.0	80.6	0.9	63.9	18.1	100.0	51
7	96.2	96.2	96.2	95.1	2.9	80.8	9.8	100.0	67
Location:									
Urban	97.8	96.8	95.5	69.3	2.3	54.1	6.9	64.5	187
Rural	99.2	97.2	98.3	77.8	0.6	66.6	11.1	69.4	292

Table 12
Tanzania
Probability Impact Calculations for
Reference Individual with Mean Characteristics
Based on Probit Estimates of the Determinants of Various Literacy Measures[†]

Literacy Measure	Household Assessment Reading	Household Assessment Writing	Household Assessment Calculation	Direct Reading – at least 2 of 6 right	Direct Reading – at least 5 of 6 right	Direct Calculation – at least 3 of 10 right	Direct Calculation – at least 6 of 10 right
Predicted Probability of Literacy for Reference Individual	.987	.975	.981	.725	.014	.592	.081
Impacts of the Following Regressors on Predicted Probability for Reference Individual							
Male (relative to female)	-0.014 (0.015)	-0.018 (0.020)	-0.033 (0.024)	-0.060 (0.047)	0.001 (0.012)	0.069 (0.061)	0.019 (0.031)
Year of schooling completed	0.003 (0.005)	0.009 (0.006)	0.011* (0.005)	0.072* (0.017)	0.002 (0.004)	0.060* (0.018)	0.012 (0.009)
Location (relative to rural): Urban	0.012 (0.012)	0.003 (0.016)	0.018 (0.016)	0.073 (0.055)	-0.016 (0.011)	0.126* (0.063)	0.040 (0.039)

* Asterisks following coefficient estimates flag coefficients that are statistically significantly different from zero at the 5 percent confidence level.
[†] See note to Table 4 for definition of probability impact.

Table 13
Tanzania
Detailed Responses to Direct Assessment Test Questions

Answer Code	Answer	Correct Answer	Share of Sample Responding with this Answer	Number of Individuals Responding with this Answer
Question1 : 105-16=				
A	89	X	67.0	321
B	91		10.4	50
C	99		12.3	59
D	111		5.6	27
None	-		4.6	22
Question 2: There are 4 rows of chairs, and 12 chairs in each row. What is the total number of chairs?				
A	12+4		22.8	109
B	12-4		10.0	48
C	12*4	X	50.5	242
D	12÷4		10.4	50
None	-		6.3	30
Question 3: Which of the following numbers is the largest?				
A	0.1		11.3	54
B	0.01		5.6	28
C	0.111	X	13.2	63
D	0.1101		65.6	314
None	-		4.2	20
Question 4: 1% of 400 is?				
A	1		6.9	33
B	4	X	34.5	165
C	40		21.1	101
D	400		30.9	148
None	-		6.7	32
Question 5: The height of a man is closest to 2?				
A	millimeters		11.1	53
B	meters	X	42.6	204
C	kilometers		19.2	92
D	centimeters		21.5	103
None	-		5.6	27
Question 6: Two cities are 12 kilometers apart. Each day, a bus makes 3 round trips between these cities. How many kilometers does the bus travel each day?				

A	72	X	21.5	103
B	36		50.5	242
C	15		14.0	67
D	4		8.1	39
None	-		5.9	28

Question 7: An island has an area of about 300 square miles. The government reports that the western 1/3 of the island is not suitable for cultivation. How many square miles of this island are suitable for cultivation?

A	50		11.7	56
B	100		23.4	112
C	150		29.2	140
D	200	X	25.9	124
None	-		9.8	47

Question 8: In an office building, each office has about 22 square miles of floor space. In this building, a square office would measure about how many meters on each side?

A	4.7	X	18.6	89
B	5.5		17.8	85
C	11		27.1	130
D	484		29.0	139
None	-		7.5	36

Question 9: If a=-3 and b=3, then $2a+b^2$?

A	0		7.7	37
B	3	X	22.8	109
C	9		30.3	145
D	12		30.5	146
None	-		8.8	42

Question 10: If $2x - 3 = 17$, then $x = ?$

A	7		19.4	93
B	10	X	19.8	95
C	14		22.3	107
D	20		23.2	111
None	-		15.2	73

QUESTIONS 11-13 REFER TO THE FOLLOWING PASSAGE:

You could smell the fish market long before you could see it. As you come closer you could hear merchants calling out about fresh catches or housewives arguing about prices. Soon you could see the market itself, brightly lit and colorful. You could see fishing boats coming in, their decks covered with silver-grey fish.

Question 11: What kind of market is described above?

A	A vegetable market		6.1	26
B	A meat market		74.5	29
C	A fish market	X	3.1	357
D	A fruit market		10.9	15
None	-		5.4	52

Question 12: Who could you see coming in?				
A	Tug boats		6.3	30
B	Row boats		64.5	309
C	Fishing boats	X	9.6	46
D	Sailboats		12.9	62
None	-		6.7	32
Question 13: What covered the decks of the boats?				
A	Rope		3.1	15
B	People		21.7	104
C	Boxes		4.4	21
D	Fish	X	64.9	311
None	-		5.9	28
QUESTIONS 14-16 REFER TO THE FOLLOWING PASSAGE: The cat brushed against the old man. He did not move. He only stood, staring up at the window of the house. The party inside looked warm and friendly, but no one noticed him. The old man walked sadly on, followed by the cat				
Question 14: What kind of animal was following the old man?				
A	Mouse		8.1	39
B	Dog		10.4	50
C	Cat	X	70.2	336
D	Bird		5.6	27
None	-		5.6	27
Question 15: What was inside the house?				
A	A party	X	56.0	268
B	Some dogs		4.4	21
C	An old lady		23.4	112
D	A meeting		9.0	43
None	-		7.3	35
Question 16: The man is described as being?				
A	Old	X	74.3	356
B	Young		4.8	23
C	Thin		9.4	45
D	Small		5.2	25
None	-		6.3	36

Table 14A Ghana GSS/OED 2003 Household and School Survey Basic Information	
Household sample design	The survey was designed to allow comparison with data collected on school facilities and cognitive abilities in the Ghana LSMS of 1988/89. The GLSMS of 1988/89 employed a self-weighting nationally representative sample derived from the frame of the 1984 Census, with households in 170 clusters. In half of those clusters data on cognitive abilities and school facilities were collected. The Ghana 2003 survey returned to those same 85 clusters areas. Within each enumeration areas new households were sampled. Thus, at least roughly speaking, the sample is nationally representative.
Direct Literacy Assessment Tests	The survey included the administration of a battery of cognitive tests. The tests in English and math were much the same as those that had been administered in the Ghana LSMS of 1988/89, but the method of determining who should completed which parts of the test was revised. A local language test was also added. All individuals in the eligible age range were administered the Raven's test of cognitive reasoning skill (thought to be a test of skill unrelated to schooling achievement). The English and math tests were adapted from the "Sabot test" developed by the Educational Testing Service in Princeton, New Jersey, and described in an appendix to Knight and Sabot (1990). In the areas of English, local language and math, most respondents were first shown a flash card of a simple sentence or simple arithmetic problem. The sentences included "They saw him and left" and "I like to eat but not to drink." The arithmetic problems included "2+3" and "3+3". If the respondent could read the word or provide the right answer to the arithmetic problem, he or she was then administered a simple test in the relevant area. The simple tests contained 8 questions each. The short tests in English and math are shown in Tables 13c and 13 d below. Individuals getting 5 or more answers correct on a simple test were given an advanced test in the relevant area. The advanced tests are included as appendices to this report. Interviewers were given the discretion to decide that it would be offensive to administer the flash card test to individuals who were almost surely literate. They seem also to have been given the discretion to administer the advanced test without first administering the simple test. Details on how the information regarding the tests was recorded in the questionnaire, and on response rates, are provided in Table 13b below. The text discusses important difficulties in analyzing the data that arise out of the coding scheme, and the rationale for the choices made in constructing literacy measures. My first direct assessment indicator in each area (English, local language and math) is meant to capture whether or not the individual did pass, or was judged by the interviewer to be able to pass (and thus was not tested), the flash card test. I set the indicator equal to one if a score is reported for either the simple or advanced English test (even if that score is zero). My second direct assessment indicator in each area is meant to capture whether or not the individual passed, or was judged by the interviewer to be able to pass, the simple reading comprehension test, with a score of five or higher. I set the

	indicator equal to one if either a score of 5 or higher is recorded on the simple test or no score was reported for the simple test but a score was reported for the advanced test.
Populations to which direct assessments were administered	According to the survey documentation, the tests were to be administered to all household members between 9 and 55 years old. In practice, it seems clear that many individuals were not tested, either because they were away or because they refused to participate. Unfortunately, the way the testing information is coded in the questionnaire leaves some ambiguity regarding who was tested in each area and who was not. Test administrators were given a listing of every household member, and this entire listing is included in the data file. The difficulties of determining who was tested and who was not are discussed in the text. In the following table, I assume that anyone for whom the Raven's test score is non-missing was indeed tested at some level in the other areas (English, local language and math). This retains about 75 percent of the individuals in the test score file who are in the appropriate age range. See text for more discussion of this issue.
Household assessment questions	Answers to the four questions: Can [name] read a newspaper in English? Can [name] write a letter in English? Can [name] read a newspaper in any Ghanaian language? Can [name] do written calculations? are used to create household assessment reading, writing and calculation indicators. A member of the household (usually the household head) responded to each of these questions on behalf of the individual in question.
Population for which household assessment questions administered	All household members 5 years old and older.
Samples examined in tables	3557 individuals between 9 and 55 years old for whom all the relevant information was available.
Definition of schooling variable	Highest grade completed in formal school. In what follows I concentrate on the following categories: 0 through 6 years of primary, and higher levels of school. I dropped the 7 observations for which the reported highest level of schooling was "Koranic", as this was a very small sample.
Weights employed	None.

Table 14 B
Ghana 2003
Questions and Response Rates among Individuals 9 to 55 Years Old for Whom Highest Grade Attained Information is Not Missing

Question	To whom addressed (in principle)	Responses	Number Providing Each Response
a2a. Simple test taken? English	All household members between 9 and 55	Nothing coded	41
		1=yes	1888
		2=no	2764
a2b Simple test taken? Local language	All household members between 9 and 55	Nothing coded	41
		1=yes	819
		2=no	3833
a2c.Simple test taken? Maths	All household members between 9 and 55	Nothing coded	41
		1=yes	2252
		2=no	2400
a2d. Simple test taken? Raven	All household members between 9 and 55	Nothing coded	41
		1=yes	3542
		2=no	1110
a3a.Easy reading score English		Nothing coded	2799
		Score	1894
a3,bEasy reading score Local language		Nothing coded	3861
		Score	832
A4.Easy Maths score		Nothing coded	2437
		Score	2256
A5.Raven Score		Nothing coded	1146
		Score	3547
A6. Reason not tested		Nothing coded	2805
		1=at work	183
		2=at school	147
		3=traveling	378
		4=too difficult/did not understand	471
		5=refusal	101
		6=illness	81
		7=age	9
		8=not eligible	46
		9=not household members	2
		10=other	470
B2aAdvanced test taken? English		Nothing coded	223
		1=yes	1029
		2=no	3441
B2bAdvanced test taken? Local language		Nothing coded	223
		1=yes	416
		2=no	4054
B2cAdvanced test taken? Maths		Nothing coded	223
		1=yes	945
		2=no	3525
B3aAdvanced reading score English		Nothing coded	3653
		Score	1040

B3bAdvanced reading score Local language		Nothing coded	4264
		Score	429
B4Advanced maths score		Nothing coded	3743
		Score	950
B5Reason not tested		Nothing coded	1200
		1=at work	171
		2=at school	132
		3=traveling	355
		4=too difficult/did not understand	1813
		5=refusal	116
		6=illness	70
		7=age	10
		8=not eligible	688
		9=not household member	5
		10=other	133

Table 14C
GHANA EDUCATION IMPACT EVALUATION, 2003

Short English Reading Test

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.

1. Who is John?

   (A) An old man
   (B) A small boy
   (C) A school teacher
   (D) A school

2. Where does John live?

   (A) In a village
   (B) In a city
   (C) In a school
   (D) In a forest

3. 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

4. How many teachers are there at John's school?

   (A) One
   (B) Three
   (C) Five
   (D) Six

5. What is John doing at school?

   (A) Helping the teacher
   (B) Talking with his friends
   (C) Learning to read
   (D) Teaching the class

6. Who is a school teacher?

   (A) John
   (B) John's father
   (C) John's brother
   (D) John's mother

7. 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

8. 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

Table 14D
GHANA EDUCATION IMPACT EVALUATION, 2003

Short Maths Test

1. 1 + 2 =	**5. 24 + 17 =**
2. 5 - 2 =	**6. 33 - 19 =**
3. 2 x 3 =	**7. 17 x 3 =**
4. 10 ÷ 5 =	**8. 41 ÷ 7 =**

Table 15A: Ghana 2003 Literacy Rate Comparisons* Females 9 to 55 Years Old									
	Household Reading English	Household Writing English	Househod Reading Local	Household Math	Passed Flash Card Test English	Passed Flash Card Test Local	Passed Flash Card Test Math	Proxy Literacy (Schooling >4)	Number of Observations
All	41.6	38.9	36.4	58.6	47.9	20.5	57.6	59.2	1891
Highest Grade Completed:									
0	0.7	0.7	1.1	4.2	4.7	0.7	6.2	0.0	449
1	11.1	3.2	4.8	30.2	28.6	9.5	50.8	0.0	63
2	10.9	6.7	8.4	41.2	26.9	5.9	44.5	0.0	119
3	24.4	17.6	19.1	48.9	45.0	17.6	58.0	0.0	131
4	35.0	33.3	22.8	69.9	58.5	16.3	74.0	100.0	123
5	43.2	36.0	36.9	65.8	58.6	24.3	73.0	100.0	111
6	52.2	47.8	40.0	80.0	63.5	25.2	79.1	100.0	115
7 or more	75.3	73.2	69.0	91.6	73.0	35.2	82.2	100.0	770
Location:									
Urban	55.8	53.6	46.8	70.6	60.4	26.4	67.1	70.8	977
Rural	26.5	23.2	25.4	45.8	34.6	14.1	47.4	46.7	914
Age group:									
9-14	40.4	34.9	28.4	63.7	59.0	20.3	71.8	45.4	507
15-24	59.3	57.4	50.6	74.3	67.9	28.2	75.5	79.8	486
25-34	31.5	30.8	33.0	50.0	33.7	15.5	44.6	59.4	406
35-44	34.5	32.9	33.5	48.4	30.0	19.4	37.7	57.7	310
45-55	32.7	29.6	33.2	39.7	27.1	15.1	33.2	44.7	199
Respondent for HouseholdAssessment:									
Self	41.6	39.6	40.8	59.6	43.3	23.1	52.2	69.0	490
Third party	41.6	38.7	34.9	58.3	49.5	19.6	59.5	55.7	1401
Other hh members with 6+ years school:									
Yes									
No	50.2	46.7	42.7	67.3	56.7	24.4	66.2	67.0	1329
	21.4	20.5	21.7	38.1	27.2	11.2	37.2	40.6	562
* See text for definitions of literacy measures and important caveats that must be place on this figures.									

Table 15B Ghana 2003 Literacy Rate Comparisons Males 9 to 55 Years Old									
	Household Reading English	Household Writing English	Household Reading Local	Household Math	Passed Flash Card Test English	Passed Flash Card Test Local	Passed Flash Card Test Math	Proxy Literacy (Schooling >4)	Number of Observations
All	56.0	53.3	48.2	69.9	62.1	28.4	72.8	68.5	1656
Highest Grade:									
0	0.7	0.4	0.7	5.8	4.7	1.1	10.6	0.0	274
1	3.4	1.7	6.9	32.8	36.2	10.3	67.2	0.0	58
2	18.1	11.1	6.9	48.6	30.6	4.2	62.5	0.0	72
3	29.5	22.9	18.1	59.0	48.6	17.1	76.2	0.0	105
4	36.3	29.7	30.8	64.8	58.2	18.7	75.8	100.0	91
5	50.0	43.5	41.7	74.1	66.7	24.1	81.5	100.0	108
6	52.9	47.9	45.4	78.2	71.4	26.9	83.2	100.0	119
7 or more	89.3	88.0	78.7	96.9	87.0	44.7	92.2	100.0	816
Location:									
Urban	68.0	65.5	55.2	81.4	74.7	33.9	81.7	78.4	846
Rural	43.6	40.6	41.0	57.9	49.0	22.7	63.6	58.1	810
Age group:									
9-14	40.3	34.0	28.5	59.9	56.1	18.6	72.7	42.9	494
15-24	65.9	65.3	56.9	79.2	74.4	34.9	83.7	83.1	496
25-34	62.5	60.3	59.2	75.7	64.4	32.6	71.5	85.0	267
35-44	60.0	58.3	53.5	67.8	53.5	28.7	61.3	71.7	230
45-55	58.6	58.0	56.4	66.3	52.5	30.9	59.1	70.2	181
Respondent for HouseholdAssessment:									
Self	59.3	57.5	57.1	70.9	57.6	31.1	66.2	77.5	604
Third party	54.2	51.0	43.2	69.4	64.7	26.9	76.6	63.3	1052
Other hh members with 6+ yrs school									
Yes	66.4	63.2	54.7	79.9	73.2	32.4	83.3	76.1	1069
No	37.1	35.3	36.5	51.8	42.1	21.3	53.7	54.5	587
* See text for definitions of literacy measures and important caveats that must be place on this figures.									

Table 15C Ghana 2003 Literacy Rate Comparisons Females 9 to 55 Years Old									
	Passed simple English test	Passed simple Local Language Test	Passed Simple Math Test						Number of Observations
All	34.9	14.2	42.9						1891
Highest Grade:									
0	1.6	0.4							449
1	7.9	4.8	3.3						63
2	6.7	0.0	14.3						119
3	24.4	7.6	15.1						131
4	37.4	7.3	38.2						123
5	45.0	13.5	52.8						111
6	35.7	11.3	56.8						115
7 or more	61.0	27.9	53.9						770
			68.4						
Location:									
Urban	48.2	19.7							977
Rural	20.7	8.3	53.3						914
			31.7						
Age group:									
9-14	36.3	11.2							507
15-24	53.5	20.6	46.0						486
25-34	25.4	12.6	60.5						406
35-44	23.5	13.9	33.3						310
45-55	23.1	11.6	33.2						199
			25.6						
Respondent for Household:									
Self	34.1	17.8							490
Third party	35.2	12.9	42.4						1401
			43.0						
Other hh members with 6+ years school:									
Yes	41.8	16.6							1329
No	18.5	8.4	49.9						562
			26.3						
* See text for definitions of literacy measures and important caveats that must be place on this figures.									

Table 15D Ghana 2003 Literacy Rate Comparisons Males 9 to 55 Years Old									
	Passed simple English test	Passed simple Local Language Test	Passed Simple Math Test						Number of Observations
All	47.8	21.8	57.0						1656
Highest Grade:									
0	2.2	0.7	5.1						274
1	10.3	3.4	15.5						58
2	11.1	1.4	31.9						72
3	23.8	2.9	42.9						105
4	33.0	9.9	50.5						91
5	38.0	9.3	59.3						108
6	50.4	16.0	65.5						119
7 or more	75.2	38.6	81.4						816
Location:									
Urban	59.9	26.2	67.3						846
Rural	35.1	17.2	46.3						810
Age group:									
9-14	33.4	9.5	47.4						494
15-24	63.1	29.8	68.5						496
25-34	48.3	23.6	58.1						267
35-44	43.9	23.9	53.0						230
45-55	49.2	28.2	54.7						181
Respondent for HouseholdAssessment:									
Self	47.4	26.5	56.6						604
Third party	48.0	19.1	57.2						1052
Other hh members with 6+ years school:									
Yes									
No	57.2	24.5	66.9						1069
	30.7	16.9	39.0						587
* See text for definitions of literacy measures and important caveats that must be place on this figures.									

Table 16A
Ghana 2003
Probability Impact Calculations for
Reference Individual with Mean Characteristics
Based on Probit Estimates of the Determinants of Various Literacy Measures[†]

	Household Reading English	Household Writing English	Household Reading Local	Household Math	Passed Flash Card Test English	Passed Flash Card Test Local	Passed Flash Card Test Math
Predicted Probability of Literacy for Reference Individual	0.377	0.322	0.314	0.687	0.516	0.170	0.671
Impacts of the Following Regressors on Predicted Probability for Reference Individual							
Male (relative to female)	0.131*	0.131*	0.070*	0.064*	0.115*	0.040*	0.126*
	(0.021)	(0.020)	(0.019)	(0.021)	(0.021)	(0.013)	(0.020)
Years of schooling completed (relative to none)							
1	0.112	-0.011	0.186*	0.171*	0.181*	0.252*	0.261*
	(0.099)	(0.119)	(0.093)	(0.036)	(0.054)	(0.079)	(0.024)
2	0.264*	0.236*	0.258*	0.254*	0.156*	0.139*	0.245*
	(0.077)	(0.095)	(0.077)	(0.023)	(0.050)	(0.071)	(0.025)
3	0.449*	0.452*	0.462*	0.297*	0.342*	0.399*	0.319*
	(0.058)	(0.074)	(0.058)	(0.018)	(0.034)	(0.063)	(0.017)
4	0.523*	0.571*	0.535*	0.329*	0.410*	0.395*	0.339*
	(0.048)	(0.057)	(0.049)	(0.015)	(0.027)	(0.064)	(0.014)
5	0.604*	0.635*	0.633*	0.340*	0.452*	0.497*	0.354*
	(0.035)	(0.045)	(0.035)	(0.014)	(0.023)	(0.058)	(0.013)
6	0.640*	0.683*	0.656*	0.362*	0.488*	0.527*	0.370*
	(0.029)	(0.036)	(0.032)	(0.014)	(0.020)	(0.056)	(0.013)
7 and up	0.927*	0.926*	0.855*	0.829*	0.797*	0.532*	0.748*
	(0.015)	(0.016)	(0.019)	(0.016)	(0.021)	(0.030)	(0.019)
Location (relative to rural):							
Urban	0.156*	0.175*	0.035	0.103*	0.151*	0.033*	0.044*
	(0.021)	(0.020)	(0.019)	(0.021)	(0.021)	(0.013)	(0.020)
Age groups (relative to 9-14):							
15-24	-0.288*	-0.228*	-0.147*	-0.196*	-0.204*	-0.057*	-0.158*
	(0.028)	(0.027)	(0.026)	(0.033)	(0.032)	(0.019)	(0.032)
25-34	-0.407*	-0.343*	-0.202*	-0.411*	-0.452*	-0.107*	-0.454*
	(0.023)	(0.023)	(0.027)	(0.042)	(0.028)	(0.018)	(0.035)
35-44	-0.334*	-0.278*	-0.170*	-0.364*	-0.461*	-0.085*	-0.491*
	(0.026)	(0.025)	(0.029)	(0.044)	(0.027)	(0.020)	(0.035)
45-55	-0.243*	-0.203*	-0.056	-0.317*	-0.407*	-0.070*	-0.448*
	(0.035)	(0.031)	(0.037)	(0.048)	(0.032)	(0.022)	(0.040)

Household report provided by (relative to other): Self	-0.013 (0.032)	-0.035 (0.029)	0.028 (0.027)	0.027 (0.032)	0.028 (0.031)	0.029 (0.019)	0.025 (0.029)
Whether household has at least one other member with 6 or more years of education (relative to none)	0.137* (0.025)	0.107* (0.024)	0.051* (0.022)	0.097* (0.025)	0.138* (0.024)	0.034* (0.016)	0.112* (0.023)

* Asterisks following coefficient estimates flag coefficients that are statistically significantly different from zero at the 5 percent confidence level.

† See note to Table 4 for definition of probability impact.

Table 16B Ghana 2003
Probability Impact Calculations for
Reference Individual with Mean Characteristics
Based on Probit Estimates of the Determinants of Various Literacy Measures[†]

	Passed Simple English Test	Passed Simple Local Language Test	Passed Simple Math Test				
Predicted Probability of Literacy for Reference Individual	0.314	0.101	0.444				
Impacts of the Following Regressors on Predicted Probability for Reference Individual							
Male (relative to female)	0.095*	0.035*	0.110*				
	(0.018)	(0.010)	(0.020)				
Years of schooling completed (relative to none)							
1	0.062	0.127	0.080				
	(0.078)	(0.077)	(0.069)				
2	0.048	-0.045	0.197*				
	(0.067)	(0.046)	(0.052)				
3	0.336*	0.174*	0.395*				
	(0.056)	(0.066)	(0.036)				
4	0.456*	0.245*	0.465*				
	(0.048)	(0.070)	(0.029)				
5	0.524*	0.314*	0.509*				
	(0.042)	(0.070)	(0.024)				
6	0.552*	0.360*	0.529*				
	(0.040)	(0.069)	(0.022)				
7 and up	0.771*	0.419*	0.770*				
	(0.024)	(0.032)	(0.020)				
Location (relative to rural):							
Urban	0.158*	0.031*	0.085*				
	(0.018)	(0.010)	(0.020)				
Age groups (relative to 9-14):							
15-24	-0.124*	-0.030*	-0.135*				
	(0.027)	(0.015)	(0.030)				
25-34	-0.297*	-0.073*	-0.327*				
	(0.021)	(0.013)	(0.028)				
35-44	-0.282*	-0.056*	-0.294*				
	(0.021)	(0.015)	(0.029)				
45-55	-0.203*	-0.035*	-0.263*				

	(0.027)	(0.017)	(0.033)				
Household report provided by (relative to other):							
Self	0.019	0.025	0.049				
	(0.027)	(0.015)	(0.028)				
Whether household has at least one other member with 6 or more years of education (relative to none)	0.097*	0.011	0.104*				
	(0.021)	(0.012)	(0.023)				

* Asterisks following coefficient estimates flag coefficients that are statistically significantly different from zero at the 5 percent confidence level.

† See note to Table 4 for definition of probability impact.

Table 17
Ghana 2003
"Passed Simple English Test" by Highest Grade Completed, in Various Subgroups

	Female		Males	
	Direct Literacy Rate	Number of Observation	Direct Literacy Rate	Numbers of Observations
0	4.7	449	4.7	274
1	28.6	63	36.2	58
2	26.9	119	30.6	72
3	45.0	131	48.6	105
4	58.5	123	58.2	91
5	58.6	111	66.7	108
6	63.5	115	71.4	119
7 and up	73.0	770	87.0	816
	Ages 9 to 15		Ages 15 to 24	
0	8.2	134	8.6	128
1	37.1	97	27.3	11
2	34.9	146	23.1	13
3	59.0	161	41.4	29
4	74.1	143	42.5	40
5	88.1	118	47.3	55
6	90.0	100	73.1	67
7 and up	96.3	81	90.4	638

	Ages 15 to 24		Ages 35 to 44		Ages 45 to 55	
	Direct Literacy Rate	Number of Observations	Direct Literacy Rate	Number of Observations	Direct Literacy Rate	Number of Observations
0	3.6	169	1.8	164	2.9	138
1	0.0	6	0.0	2	0.0	5
2	0.0	13	0.0	10	0.0	10
3	0.0	16	10.0	20	9.1	11
4	10.0	10	0.0	11	9.1	11
5	19.0	21	11.1	18	14.3	7
6	43.8	32	10.0	20	20.0	15
7 and up	70.1	405	70.2	295	76.0	183

	Rural		Urban	
	Direct Literacy Rate	Number of Observations	Direct Literacy Rate	Number of Observations
0	3.9	519	6.9	204
1	24.3	70	43.1	51
2	22.3	103	35.2	88
3	33.3	123	61.1	113
4	56.1	107	60.7	107
5	54.5	101	69.5	118
5	57.6	99	74.8	135
7 and up	74.3	591	83.7	995

Appendix 1

Bangladesh Survey
LITERACY TEST

Self Assessment:

Can you read a letter? আপনি কি পড়তে পারেন ?	**Yes** হ্যাঁ	**No** না
Can you write a letter? আপনি কি লিখতে পারেন ?	**Yes** হ্যাঁ	**No** না

READING, Level 1

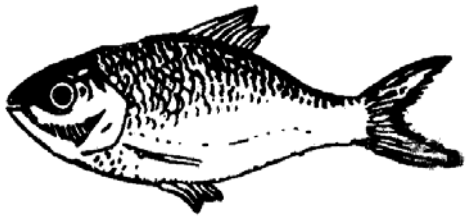
1. **Show me the letter** অ
2. **Show me the letter** চ
3. **Show me the letter** এ
4. **Show me the letter** ফ
5. **Show me the letter** ম
6. **Show me the letter** দ
7. **Show me the letter** খ
8. **Show me the letter** ব

অ	আ	ই	উ	এ	
ক	গ	চ	জ	ট	
	ঈ	ঊ	ঋ	এ	ঔ
	ঙ	ফ	ণ	ঞ	ষ
	প	ক	ম	ঠ	র
	ত	ছ	দ	ছ	স
	থ	খ	ঝ	ল	শ
	ঈ	ন	ব	ড	ভ

At least 7 items have to be answered correctly to pass this level.

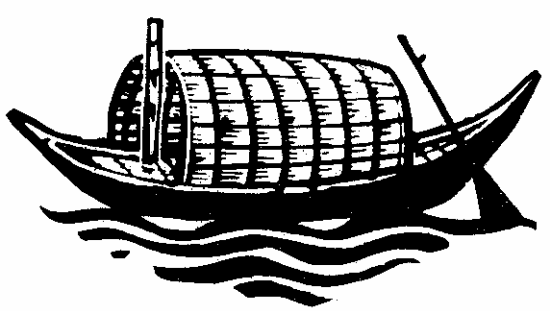
READING, Level 2

9. Underline the word that tells what is in this picture.



পাখি মাছ কলা ফুল চাবি

10. Underline the word that tells what is in this picture.



পানি গাভী নদী নৌকা ছাতা

11. Underline the word পানি — জমি মাতা দাদা পানি মাছ

12. Underline the word কলা — মই কলা নারী ছাতা পাখি

13. Underline the word জামা — জামা চাকা খুকু বসা কথা

14. Underline the word নদী নালা — নরনারী মহাজন গাছপালা নদী নালা

15. Underline the word তরকারী — জাতিভেদ জমিজমা সখিপুর তরকারী

16. Underline the word গজারিয়া — কাজিবাড়ী হাসিখুশী গজারিয়া সুখদুঃখ

At least 6 items have to be answered correctly to pass this level.

READING, Level 3

Father said that he will sow masoor (lentil) and mustard on one side of this land and cauliflower, potatoes, tomatoes and winter vegetables on the other side. We get our food for the whole year from this land. The smile on my mother's face resembles the grown harvest. What a pleasure to see the smile on my mother's face!

বাবা বলেছেন, এ জমিতে তিনি কয়েকদিন পরে একদিকে বুনবেন মসুর আর সরিষা। অন্যদিকে লাগাবেন কপি, আলু, টমেটো আর শীতের শাকসবজি। এ জমি থেকেই আমরা সারা বছরের খাবার পাই। পাকা ফসলের মত আমার মায়ের মুখে হাসি লেগে থাকে। মায়ের মুখের হাসি আমার কি যে ভাল লাগে।

17. Who first said the above words?

উপরের কথাগুলি প্রথম কে বলেছেন ?

18. Name two vegetables mentioned here.

যে সব শাকসবজি লাগাবেন তাদের মধ্যে দু'টির নাম বলুন ?

19. How many days worth of food will they get from the land?

এ জমি থেকে তারা কতদিনের খাবার পায় ?

20. Whose smile is a source of pleasure?

কার মুখের হাসি দেখতে ভাল লাগে ?

21. What does that smile resemble?

মায়ের মুখে কিসের মত হাসি লেগে থাকে ?

At least 3 of the items have to be answered correctly to pass this level.

WRITING, Level 1

1. **Write the letter** আ — লিখুন আ
2. **Write the letter** প — লিখুন প
3. **Write the letter** উ — লিখুন উ
4. **Write the letter** ব — লিখুন ব
5. **Write the letter** থ — লিখুন থ
6. **Write the letter** ট — লিখুন ট
7. **Write the letter** চ — লিখুন চ
8. **Write the letter** ম — লিখুন ম

At least 6 items had to be answered correctly to pass this level (Those who failed were placed on Level 1).

WRITING, Level 2

LISTEN CAREFULLY AND WRITE DOWN THE WORD THAT YOU HEAR.

9. **Write the word *"ball"* .** লিখুন বল
10. **Write the word *"mango"*.** লিখুন আম
11. **Write the word *"key"* .** লিখুন চাবি

Write these words:

12. ***I will go to the market.*** আমি বাজারে যাব
13. ***I don't have a shirt.*** আমার জামা নাই
14. ***There will be good crops this year.*** এবার ভাল ফসল হবে

At least 4 items have to be answered correctly to pass this level.

WRITING, Level 3:

Look at this picture. Write a short story of what you see in the picture. (A minimum of 12 words of understandable text had to be written within a 6 minute period).

MAT

HEMATICS ORAL, Level 1:

1. You have 6 cows. Your brother has 3 cows. How many cows do you have in all?

আপনার ৬টি গরু আছে, আর আপনার ভাইয়ের আছে ৩টি গরু। দু'জনের মোট কয়টি গরু আছে ?

2. You have 12 goats. Your uncle has 6 goats. How many goats do you have in all?

আপনার ১২টি ছাগল আছে এবং আপনার চাচার ৬টি ছাগল আছে। আপনাদের দু'জনের মোট কতটি ছাগল আছে ?

3. **You have 15 chickens. You sell 6 of them. How many chickens have you left?**

আপনার ১৫টি মুরগীর বাচ্চা আছে, এর মধ্যে থেকে ৬টি মুরগীর বাচ্চা বিক্রী করে দিলে কতটি থাকবে ?

4. **You have 8 five taka notes. How much money does you have?**

আপনার কাছে ৮টি ৫ টাকার নোট আছে। আপনার কাছে মোট কত টাকা আছে ?

5. **Four eggs cost 12 taka. How much does one egg cost?**

৪টি ডিমের দাম ১২ টাকা হলে, ১ টি ডিমের দাম কত ?

At least 4 items had to be answered correctly to pass this level.

***MATHEMATICS ORAL, Level 2*:**

6. **Jamil went to market with two 100 taka notes and ten 5 taka notes. How much does he have in all?**

জামিল ২টি ১০০ টাকার নোট ও ১০টি ৫ টাকার নোট নিয়ে বাজারে গেল। সে মোট কত টাকা নিয়ে বাজারে গেল ?

7. **You save 20 taka each month. How much will you have saved after 6 months?**

আপনি প্রতি মাসে ২০ টাকা করে জমান। ৬ মাসে আপনার মোট কত টাকা জমবে ?

8. **Selim has 250 taka. Chickens cost 60 taka each. How many chickens can he buy? What change will he get?**

সেলিম ২৫০ টাকা নিয়ে বাজারে গেল মুরগী কিনতে। প্রতিটি মুরগীর দাম ৬০ টাকা হলে সে কতটি মুরগী কিনতে পারবে এবং তার কাছে আর কত টাকা থাকবে ?

9. **If 100 taka earns 5 taka interest per month. How much principle & interest will you have after one**

১০০ টাকার ১ মাসের সুদ ৫ টাকা হলে ১ বছর পর সুদ আসলে মোট কত টাকা দিতে হবে ?

At least 3 of these questions had to be answered correctly to pass this level.

MATHEMATICS WRITTEN, Level 1

1. **Show me the number eight:** 1 2 3 4 5 6 7 8 9
   ১ ২ ৩ ৪ ৫ ৬ ৭ ৮ ৯

2. **Show me the number five:** 1 2 3 4 5 6 7 8 9
   ১ ২ ৩ ৪ ৫ ৬ ৭ ৮ ৯

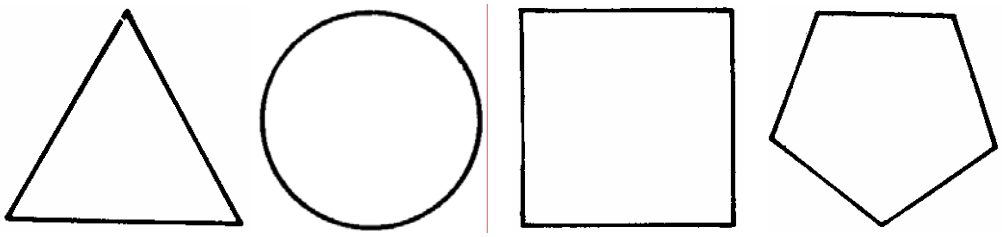
3. **Show me the number thirty-nine:** 3 9 93 30 39
   ৩ ৯ ৯৩ ৩০ ৩৯

4. **Show me the number fifty-eight:** 50 58 8 85 78
   ৫০ ৫৮ ৮ ৮৫ ৭৮

5. **Write the number six:** ৬ অংকটি লিখুন

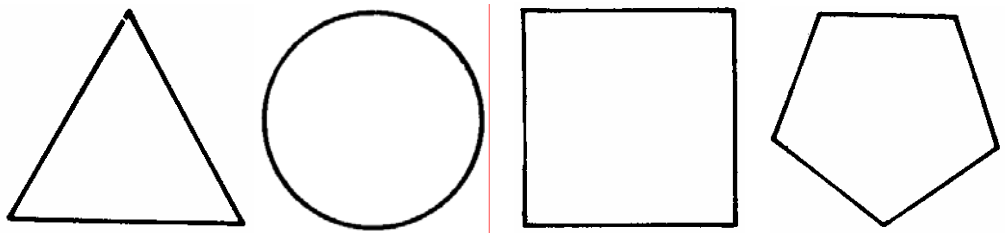
6. **Write the number nine:** ৯ অংকটি লিখুন

7. **Pick out the circle:**
   কোনটি গোলাকার দেখান



8. Pick out the square:

কোনটি চারকোনা দেখান



At least 6 questions had to be answered correctly to pass this level.

MATHEMATICS WRITTEN, Level 2

9. Add:

12
+26

১২

যোগ করুন ঃ ২৬

10. Subtract:

17
- 5

১৭

বিয়োগ করুন ঃ ৫

11. Multiply:

7
x9

৭

গুণ করুনঃ ৯

12. Divide 32 by 8:

৩২ কে ৮ দিয়ে ভাগ করুন

At least 3 of the questions had to be answered correctly to pass this level.

MATHEMATICS WRITTEN, Level 3

13. 365 equals one year. Last year, we had rain for 123 days. How many days didn't it rain?

৩৬৫ দিনে ১ বছর। গত বছর ১২৩ দিন বৃষ্টি হয়েছিল। তাহলে কতদিন বৃষ্টি হয়নি ?

14. Shafique has 6 cows. Rafique has twice as many as Shafique. How many cows do they have together?

শফিকের ৬টি গরু আছে। রফিকের, শফিকের চেয়ে দ্বিগুণ গরু আছে। শফিক ও রফিকের দু'জনের মোট কতটি গরু আছে ?

15. Rahima has 32 mangoes. She divided them equally among her 4 children. How many did each child receive?

রহিমার ৩২টি আম আছে। তিনি আমগুলো তাঁর ৪টি ছেলেমেয়ের মধ্যে সমান ভাগে ভাগ করে দিলেন। ছেলে মেয়েদের প্রত্যেকে কতটি করে আম পেল ?

16. Bangladesh has 18 districts and each district has 7 thana's. How many thana's are in the 18 districts?

নিচে বছর ভিত্তিক ২টি গ্রামের খানা সংখ্যা দেয়া আছে। ১৯৯১ সালে ২টি গ্রামের খানা সংখ্যা কত ছিল ?

At least 3 of the questions had to be answered correctly to pass this level.

Appendix 2

GHANA STATISTICAL SERVICE

Advanced English Test

Ghana Education Impact Evaluation

INSTRUCTIONS

This test consists of 29 multiple-choice questions. For each question there is only one correct answer. You write the letter (A, B, C or D) corresponding to the correct answer for each question on your answer sheet. The maximum time allowed for this test is 20 minutes.

January-March 2003

Directions: For questions 1-9, read the passages below. Each passage is followed by questions. Choose the correct answer to each question and mark the letter of that answer on your answer sheet.

The Herring Gull is especially good at seizing food from other birds. It is about twenty-four inches long, and it is the gull that you most often see at the beach. It will often chase a bird that is carrying a fish or a stolen egg home to eat. The Herring Gull keeps attacking the other bird until it drops the egg or the fish. Of course the egg will break if it hits the ground. But Herring Gulls are so fast and agile they can sometimes catch an egg in mid-air.

1. What is a Herring Gull?

   (A) A bird
   (B) A fish
   (C) An egg
   (D) A beach

2. Which of the following is the best title for this passage?

   (A) How Herring Gulls Get Food
   (B) Catching Eggs
   (C) How Herring Gulls Fly Faster than Other Birds
   (D) Eating Habits of Birds

3. How long is a Herring Gull?

   (A) 12 inches
   (B) 18 inches
   (C) 24 inches
   (D) 32 inches

You could smell the fish market long before you could see it. As you came closer you could hear merchants calling out about fresh catches or housewives arguing about prices. Soon you could see the market itself, brightly lit and colourful. You could see fishing boats coming in, their decks covered with silver-grey fish.

4. What kind of a market is described above?

   (A) A vegetable market

   (B) A meat market

   (C) A fish market

   (D) A fruit market

5. What could you see coming in?

   (A) Tug boats

   (B) Rowboats

   (C) Fishing boats

   (D) Sailboats

6. What covered the decks of the boats?

   (A) Rope

   (B) People

   (C) Boxes

   (D) Fish

The cat brushed against the old man. He did not move. He only stood, staring in the window of the house. The party inside looked warm and friendly, but no one noticed him. The old man walked sadly on, followed by the cat.

7. What kind of animal was with the old man?

(A) Mouse
(B) Dog
(C) Cat
(D) Bird

8. What was inside the house?

(A) A party
(B) Some dogs
(C) An old lady
(D) A meeting

9. The man is described as being

(A) Old
(B) Young
(C) Thin
(D) Small

Directions: For questions 10-15, read the passage below. Each line of the passage has a number. In each line, there is a box with four possible choices. Pick the choice that best completes the sentence in each numbered line. Mark the letter (A,B,C, or D) of the choice on your answer sheet.

10. Sound is something we

(A)	hears.
(B)	hearing.
(C)	heard.
(D)	hear.

It comes to your

11.

(A)	Eyes
(B)	nose
(C)	ears
(D)	mouth

in different ways. It might be pleasant,

12. like the voice of a friend,

(A)	when
(B)	as
(C)	or
(D)	since

unpleasant, like the <u>yelp</u> of a

13. dog that has been struck by a

(A)	horn.
(B)	car.
(C)	road.
(D)	bridge.

Some sounds are loud,

14. and some are soft; some are high, and some are

(A)	full.
(B)	low.
(C)	quite.
(D)	big.

Sound is

15. very

(A)	importance
(B)	importantly
(C)	important
(D)	import

to us because it is the basic means of

communication.

Questions 16-18 are also about the group of sentences on the previous page. Choose the best answer for each of these questions and mark it on your answer sheet.

16. What does yelp in line 12 means?

(A) noise
(B) motion
(C) place
(D) piece

17. Which of the phrases below is another example of a pleasant sound, similar to the phrase in the sentence that begins in line 12, ‘like the voice of a friend’?

(A) Like the hiss of a snake
(B) Like the honk of a horn
(C) Like the rumble of thunder
(D) Like the song of a bird

18. Which sentence below has almost the same meaning as the sentence that begins in line 14?

(A) It is meaningful to communicate with sounds
(B) The main way we communicate is with sounds
(C) The meaning of sound is basic to communication
(D) In order to communicate, we need basic sounds

Directions: For question 19-29, read the passage below. Each line of the passage has a number. In each line, there is a box with four possible choices. Pick the choice that best completes the sentence in each numbered line. Mark the letter (A, B, C, or D) of the choice on your answer sheet.

19. In the late eighteen century England had all the things necessary for the growth of

(A) industry.
(B) industrial.
(C) industrially
(D) industrialize.

20. She possessed money to invest, an ample labor supply

(A) too
(B) and
(C) however
(D) although

essential natural resources.

21. Since the sixteenth century, wealth had been pouring into England from colonies in America and

(A) trade
(B) traded
(C) trades
(D) trading

22. posts in Asia. Her large merchant fleet brought her

(A) an example
(B) an abundance
(C) a contribution
(D) a distinction

of raw materials more than could

23. be used

(A) on
(B) by
(C) over
(D) against

the old system of manufacture. This surplus and the growing demand for goods encouraged

24. the development of new and faster methods of manufacture. Another favorable factor in England

(A) is
(B) was
(C) has been
(D) has to be

25. the supply of cheap labor.

(A) Sailors
(B) Diplomats
(C) Beggars
(D) Workers

could move about freely and employers had greater liberty in

26. deciding what to make and how to make it. Finally the presence of large amounts of coal

(A) made
(B) are made
(C) is making
(D) will make

possible

27.

(A) predictable
(B) illustrative
(C) extensive
(D) individual

use of steam power, and the plentiful supply of iron encouraged the manufacture of tools and machinery.

28. No other country in Europe had

(A) so
(B) too
(C) such
(D) much

favorable conditions for a vast increase in production.

29.

(A) As a result
(B) Likewise
(C) In addition
(D) On the other hand

the Industrial Revolution had its beginning in England.

Appendix 3

GHANA STATISTICAL SERVICE

Advanced maths test

Ghana Education Impact Evaluation

TRUCTIONS

This test consists of 36 multiple choice questions. For each question there is only one correct answer. You write the letter (A, B, C or D) corresponding to the correct answer for each question on your answer sheet. The maximum time allowed for this test is 30 minutes.

January-March 2003

1. $\begin{array}{r} 105 \\ -\ 16 \\ \hline \end{array}$

(A) 89

(B) 91

(C) 99

(D) 111

---

2. $4\overline{)2836}$

(A) 79

(B) 201 r 2

(C) 701 r 2

(D) 709

---

3. There are 4 rows of chairs and 12 chairs in each row. How do you find out the total number of chairs?

(A) 12 + 4

(B) 12 - 4

(C) 12 x 4

(D) 12 ÷ 4

```
 ---10 km ---
|----------------|------------------------|
A                B                        C
```

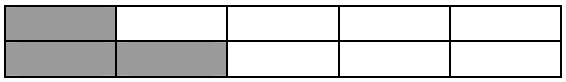
4. In the figure above, if the distance from A to C is 30 kilometres, what is the distance in kilometres from B to C?

(A) 30 + 10

(B) 30 - 10

(C) 30 x 10

(D) 30 ÷ 10



5. What part of the figure above is dark?

(A) $\frac{1}{3}$

(B) $\frac{3}{5}$

(C) $\frac{3}{7}$

(D) $\frac{3}{10}$

---

6. Which is NOT equal to $\frac{4}{10}$

(A) $\frac{1}{4}$

(B) $\frac{2}{5}$

(C) $\frac{10}{25}$

(D) $\frac{20}{50}$

---

7. Which is between and $\frac{4}{3}$

(A) 1

(B) 3

$\frac{3}{4}$

(C) $\frac{1}{2}$

(D) $\frac{3}{2}$

---

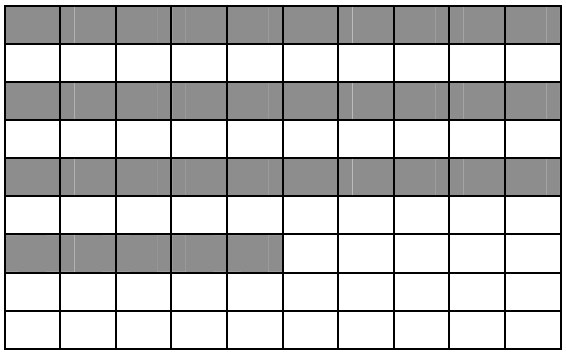
8. $\frac{1}{2} + \frac{1}{3} =$

(A) $\frac{1}{5}$

(B) $\frac{2}{6}$

(C) $\frac{2}{5}$

(D) $\frac{5}{6}$



9. If the large square above represents one unit, which decimal shows the amount that is dark?

(A) 0.35

(B) 3.5

(C) 30.5

(D) 35.100

_____

10. 0.25 =

A) $2\frac{1}{2}$

(B) $\frac{2}{5}$

(C) $\frac{1}{2}$

(D) $\frac{1}{4}$

_____

_____

11. 2 – 0.1 =

(A) 1

(B) 1.9

(C) 2.1

(D) 2.9

---

12. 0.6 x 10 =

(A) 0.06

(B) 0.60

(C) 6

(D) 60

---

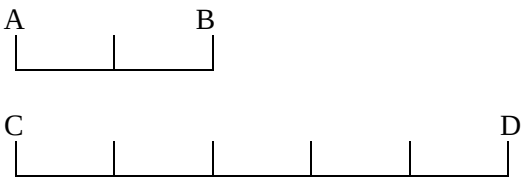
13. Which of the following numbers is the LARGEST?

(A) 0.1

(B) 0.01

(C) 0.111

(D) 0.1101



14. In the drawing above, if the distance from A to B is 6 metres, then the distance from C to D in metres is

(A) 5

(B) 12

(C) 15

(D) 30

---

15. 1% of 400 is

(A) 1

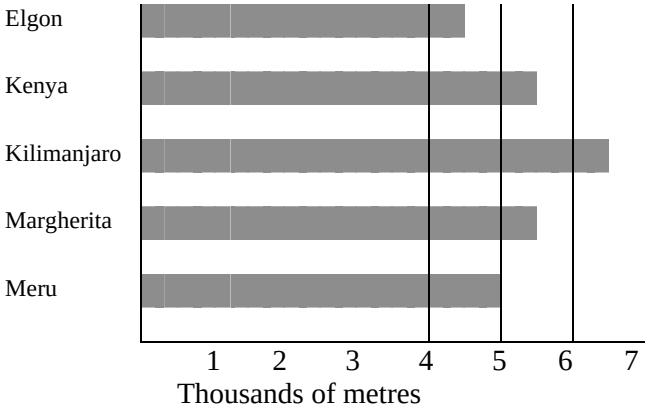
(B) 4

(C) 40

(D) 400

Items 16-17 refer to the following graph:

Heights of Five East African Mountain Peaks in Metres



16. According to the graph, the height of Mt. Kilimanjaro, in meters, is about

(A) 6,000
(B) 6,005
(C) 6,050
(D) 6,500

---

17. Which of the two peaks are most nearly the same height?

(A) Kilimanjaro and Kenya
(B) Kenya and Margherita
(C) Meru and Elgon
(D) Margherita and Meru

---

18. The height of a man is closest to 2

(A) Millimetres
(B) Metres
(C) Kilometres
(D) Centimetres

---

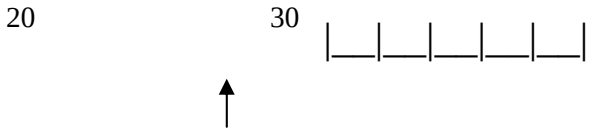
19. There are twelve inches in one foot. How many square inches are there in a square foot?

(A) 12

(B) 36

(C) 48

(D) 144



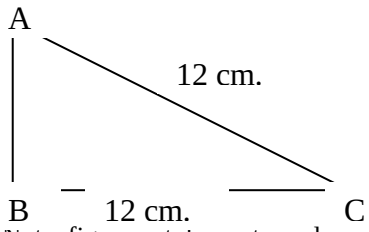
20. In the figure above the arrow points to

(A) $20\frac{4}{5}$

(B) $20\frac{5}{6}$

(C) 24

(D) 28



Note: figure not drawn to scale

21. If the perimeter of the triangle ABC is 30 centimetres, what is the length, in centimetres of side AB?

(A) $2\frac{1}{2}$

(B) 3

(C) 6

(D) 18

---

22. Two cities are 12 kilometres apart. Each day, a bus makes 3 round trips between these cities. How many kilometres does the bus travel each day?

(A) 72

(B) 36

(C) 1

(D) 4

---

23. A meal costs 1500 Cedis. If a 10% service charge is to be added to the bill, what would the total charge be?

(A) 1510 Cedis

(B) 1600 Cedis

(C) 1650 Cedis

(D) 2500 Cedis

---

24. An island has an area of about 300 square miles. The government reports that one third of the island is not suitable for cultivation. About how many square miles of this island are suitable for cultivation?

(A) 50

(B) 100

(C) 150

(D) 200

	Highest	Lowest
Elderet	23.6 °	9.5 °
Magadi	34.9 °	23.1 °
Nakura	26.4 °	10.1 °
Narok	24.4 °	8.3 °

25. The chart above shows the average (mean) high and low temperatures for four cities in a certain year. In which of the cities was there the greatest difference between the average high and the average low?

(A) Eldoret

(B) Magadi

(C) Nakura

(D) Narok

---

26. In an office building, each office has about 22 square metres of floor space. In this building, a square office would measure about how many metres on each side?

(A) 4.7

(B) 5.5

(C) 11

(D) 484

---

27. One number is 3 more than twice another. If x represents the smaller number which of the following represents the larger number?

(A) $2x + 3$

(B) $5x$

(C) $3(2x)$

(D) $2x-3$

---

28. if $a = -3$ and $b = 3$, then $2a + b^2 =$

(A) 7

(B) 3

(C) 9

(D) 12

---

29. If $2x – 3 = 17$, then $x =$

(A) 7

(B) 10

(C) 14

(D) 20

---

30. $x + \frac{1}{2}$

(A) $\frac{1}{x+2}$

(B) $\frac{x+1}{x+2}$

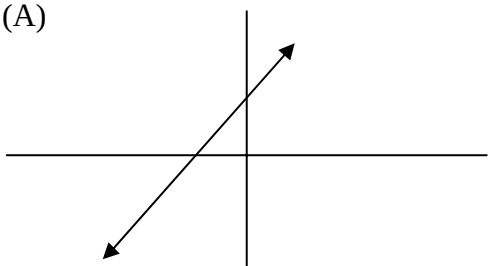
(C) $\frac{2x+1}{2}$

(D) $\frac{x+1}{2}$

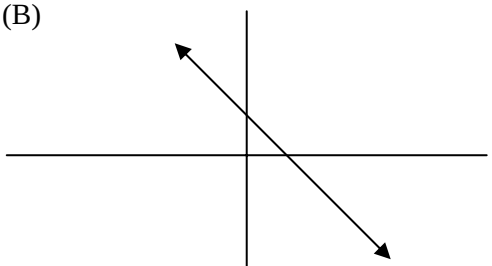
---

31. Which of the following shows the graph of $x - y = 2$?

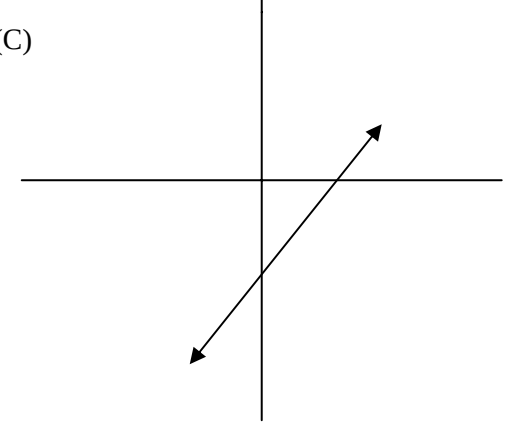
(A)



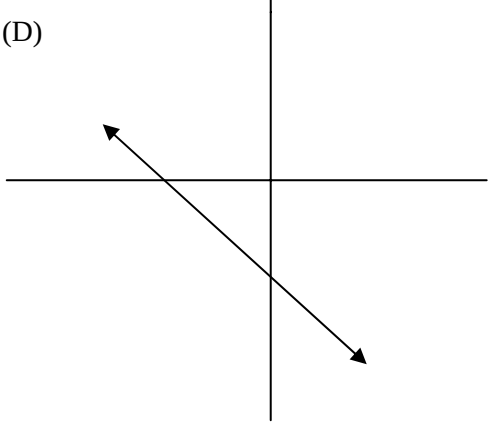
(B)

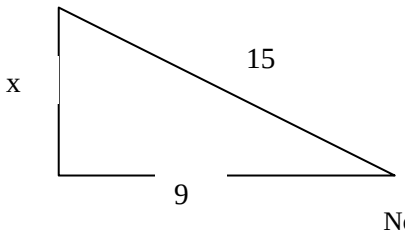


(C)



(D)





Note: Figure not drawn to scale

32. In the triangle above, x =

(A) 6

(B) 12

(C) 24

(D) $\sqrt{306}$

---

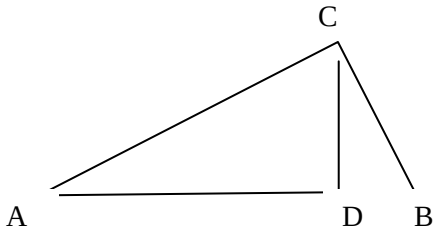
33. In quadrilateral ABCD, angle A = 60 °, and the other three angles are equal. What is the degree measure of angle B ?

(A) 220 °

(B) 100 °

(C) 60 °

(D) 40



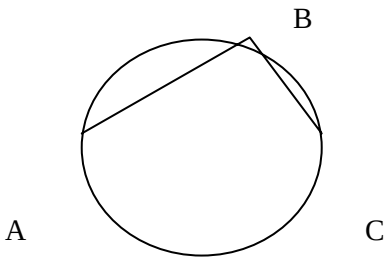
3 .. In triangle ABC above, CD is an altitude to AB, and angle ACB is a right angle. Which of the triangles shown must be similar?

(A) none

(B) triangle ACD and triangle CBD only

(C) triangle ABC and triangle ACD only

(D) triangle ABC, triangle ACD, and triangle CBD



35. In the figure above, the angle ABC is a right angle. If the centre of the circle is called Q, what can be said about the location of Q?

(A) Q is inside triangle ABC

(B) Q is outside triangle ABC

(C) Q is on $\overline{AC}$

(D) The location of Q depends on the lengths of $\overline{AB}$ and $\overline{BC}$

---

36. Which CANNOT be the intersection of 3 planes?

(A) 1 point

(B) 1 line

(C) 3 concurrent lines

(D) 3 parallel lines