

EFA 2000 Assessment Surveys Report

**ASSESSING BASIC LEARNING COMPETENCIES
AMONG YOUTH AND YOUNG ADULTS
IN DEVELOPING COUNTRIES:**

**Analytic Survey Framework
and Implementation Guidelines**

Final Report Based on
Expert Workshop for EFA2000
April 12-14 1999, ILI-UNESCO, Philadelphia

Co-organized by



INTERNATIONAL LITERACY INSTITUTE

Produced and printed by ILI, Philadelphia, September 1999

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

EXECUTIVE SUMMARY

1. INTRODUCTION.....5

2. POLICY FRAMEWORK.....5

 2.1 PRIORITY POPULATIONS..... 5

 2.2 POLICY AND ASSESSMENT NEEDS..... 6

 2.3 CHALLENGES FOR POLICY-DRIVEN ASSESSMENTS 8

 2.4 WHAT MIGHT POLICY MAKERS WISH TO KNOW ABOUT BLCs? 9

3. DEFINING BASIC LEARNING COMPETENCIES (BLCs)9

 3.1 CHALLENGES IN DEFINING BLCs..... 10

 3.2 DIRECT AND INDIRECT MEASURES 10

 3.3 CONCEPTUAL OVERVIEW OF ASSESSMENT MATRIX 11

 3.4 DEFINING LITERACY (READING AND WRITING)..... 12

 3.5 DEFINING NUMERACY 15

 3.6 DEFINING LIFE SKILLS 16

4. SAMPLING AND SURVEY DESIGN 17

 4.1 PROPOSED LOW-COST BLC HOUSEHOLD SURVEY DESIGN 17

 4.2 SOME ADVANTAGES OF THIS HOUSEHOLD SURVEY DESIGN 18

 4.3 PRINCIPLES FOR EFFECTIVE HOUSEHOLD SURVEY DESIGNS 19

 4.4 COST CONSIDERATIONS 21

5. MEASURING BLCs: AN ASSESSMENT MODEL..... 21

 5.1 LEVELS OF BLC SKILLS 22

 5.2 READING ASSESSMENT 22

 5.3 WRITING ASSESSMENT..... 24

 5.4 NUMERACY ASSESSMENT..... 25

 5.5 LIFE SKILLS ASSESSMENT..... 26

6. IMPLEMENTATION AND CAPACITY BUILDING..... 28

 6.1 DESIGN DECISIONS AND OPTIONS FOR BLC SURVEYS 28

 6.2 HUMAN RESOURCES: RECRUITING AND TRAINING PERSONNEL..... 29

 6.3 TEST DEVELOPMENT, PRETESTING AND PILOT STUDIES..... 30

 6.4 ENSURING THE QUALITY OF DATA COLLECTION..... 32

 6.5 DATA PROCESSING AND ANALYSIS 32

 6.6 REPORTING AND DISSEMINATION 33

7. CONCLUSIONS AND NEXT STEPS..... 33

TABLES

REFERENCES

ENDNOTES

ANNEXES

ACKNOWLEDGEMENTS

This report is an outcome of a collaboration between the International Literacy Institute and UNESCO. It began with a first expert seminar at UNESCO-Paris in June 1998 entitled "Literacy Assessment in Out-of-School Youth and Adults," and from which a final report was published in December 1998. The present report is based on a follow-up workshop which was held in Philadelphia in April 1999.

The overall goal of this activity is three-fold: to try to improve our understanding of the nature of basic learning competencies (BLCs) within and across societies; to assist governments and multinational agencies in obtaining reliable and useful information in this domain in order to improve in-school and out-of-school learning; and to assist schools and programs that provide instruction to better understand the effectiveness of their activities.

This project has been undertaken in the lead up to the EFA2000 activities in the year 2000. Original funding was provided to UNESCO by the IPIEO program funded by the World Bank to improve international data gathering. It also received support from within UNESCO from the EFA, MLA, and ED/BAS programs. Within UNESCO, we would especially like to thank Vina Chinapah, Dieter Berstecher, Aicha Diallo Bah and Abdulwahid Yousif for their support and encouragement.

The participants in the present workshop are listed in Annex 1 of this report, and we are most grateful to them for their insights and critical comments during the workshop, as well as to various drafts which followed the workshop. Given the complexity of the task of trying to balance between the varied goals of policy intent, statistical reliability, realistic scope of future studies, and limits of time for this work, we hope that our co-participants and the readers of this report will understand why we could not cover every topic requested, or respond affirmatively to every interesting idea. While we benefited from many comments, all remaining errors are those of the principal authors: Dan Wagner, John Sabatini, and Iddo Gal.

EXECUTIVE SUMMARY

The World Conference on Education for All (WCEFA) in 1990 at Jomtien, Thailand, included a number of educational targets related to out-of-school youth and adults, including: (1) to reduce the number of adult illiterates to half of the 1990 level by the year 2000; and (2) to improve learning achievement to an agreed percentage of an appropriate age cohort. WCEFA also stressed the need to monitor and evaluate the performance of individual learners as well as the delivery mechanisms and outcomes of literacy and other non-formal education programmes. At the Mid-Decade Review meeting on EFA in Amman (in June 1996), the international community further called for efforts at both international and national levels to adopt new techniques and strategies to collect and analyze meaningful data to monitor progress towards the Jomtien goals. In the year 2000, UNESCO and other agencies will gather together worldwide data on education in order to take stock of the worldwide progress toward the WCEFA goals, a decade after Jomtien.

At a prior expert seminar held at UNESCO-Paris in June 1998, the focus was on describing and analyzing the need for better assessment and monitoring of basic learning competencies (BLCs), from the basic skills (reading, writing, math) to 'basic life skills', as well as describing methodologies for assessments at the national survey level and at the program level.

Following on the June 1998 seminar, the present workshop (organized jointly by the International Literacy Institute and UNESCO, and held on April 12-14 1999, in Philadelphia), focused on the creation of a model for BLC assessment, especially for out-of-school youth and young adults. The present report provides both an analytical framework and implementation guidelines, with five main dimensions: Policy Framework; Defining Basic Learning Competencies (BLCs); Sampling and Survey Design; Measuring BLCs: An Assessment Model; and Implementation and Capacity Building.¹

This report describes a low-cost and adaptable methodology for the assessment of basic learning competencies in diverse cultural contexts. Next steps would involve a set of initial surveys, including assessment instrumentation, for use either at the national level or at a program level. This report, conducted in light of the 1990 Jomtien Conference and in anticipation of the EFA2000 evaluation activities, is a reflection of the increased concern about the status of basic learning competencies in many nations, as well as about the effectiveness of education program (formal and non-formal) at all levels. Through the use of a BLC assessment framework, such as described in this report, information collection for policy formulation and planning can be much improved in the coming years.

1. INTRODUCTION

The World Conference on Education for All (WCEFA), held in 1990 in Jomtien, Thailand, has highlighted the need to improve learning opportunities and learning achievement in all countries. Article 1 of the WCEFA Declaration states that "basic learning needs" or competencies (BLCs),

"comprise both essential learning tools (such as literacy, oral expression, numeracy and problem solving) and the basic learning content (such as knowledge, skills, values, and attitudes) required by human beings..." (Unesco, 1990, p. 43)

In Jomtien, a number of educational targets related to youth and adults were agreed upon, including reducing by 50% the number of adult illiterates by the year 2000 and improving learning achievement to an agreed percentage of an appropriate age cohort. At the Mid-decade Review meeting on Education For All in Amman (in June 1996; see UNESCO, 1996), the international community further called for efforts at both international and national levels to adopt new techniques and strategies to collect and analyze meaningful data to monitor progress towards the Jomtien goals.

Although many countries have been actively striving to meet basic learning needs for all, current national and international capacities remain inadequate for assessing and monitoring the acquisition or status of literacy, numeracy, and various key life-skills which are part of the desired BLCs of youth and young adults.

This report, based on the Expert Workshop on BLC Assessment held in Philadelphia in April 1999, provides an overview and programmatic synthesis of the issues facing countries that want to develop credible, feasible, and sustainable assessments of basic learning competencies. The report focuses primarily on youth and young adult populations (with an emphasis on those out-of-school), in order to provide a sound basis for social policy setting and inform educational planning.

2. POLICY FRAMEWORK

2.1 Priority populations

The population priority of this report includes out-of-school youth and young adults in the age range of 12-20 years - those who have never attended school or who dropped out before the end of compulsory education. Age range priorities can vary across countries and agencies in the area of youth and young adults. The lower point of the age range may be 12 years in some countries or 14 years in other countries, depending on the age at which compulsory schooling ends in a given country. At the upper bound, some countries may choose 16 or 18 years of age rather than 20.

For the purposes of the present report, the age range was determined in light of both general policy goals and expected logistical and funding-related constraints on surveys, and as a function of the technical options available to test construction (where too young an age can make certain test items difficult to administer).

Further, this report focuses on "disadvantaged" out-of-school youth and young adults - those who have had limited or no opportunities to access or benefit from sustained learning in formal school settings, due to their gender, social, economic, geographical, ethnic or political circumstances. This target group is comprised for the most part of young people (the majority of which are often girls or young women, especially in rural areas). They may or may not be wage earners or capable of participating in the economy and in community activities, but their basic or functional literacy, numeracy, and problem-solving skills, as well as "life skills" (related to managing personal health, nutrition, and workplace contexts) are likely to be lower than desired by national policy makers and/or what is needed to assure economic survival.

Despite their disadvantaged situation, such youth and young adults have important economic and civic roles to play in their countries. Skill deficiencies may influence their readiness to cope with a range of everyday tasks and situations, and in turn result in pressures on (and extra costs for) medical, welfare, employment-related, and community services. Hence, the monitoring of BLCs will be beneficial to many countries that seek to review policies and programs in human development and education.

While a majority of these young people have had relatively little schooling in most developing countries, they are still young enough to be able to engage and take advantage of new or additional educational opportunities in either formal or non-formal educational settings. Indeed, they may be able to do so more easily than "older" adults who already have families or who are less likely to be able to enter the new and dynamic aspects of the global economy. Hence, a policy focus on out of school youth and young adults is especially important, as this key group may be able to contribute most to new social and economic dynamics.

2.2 Policy and assessment needs

Significant percentages of children, youth and young adults in many developing countries have had little or no formal education (UNESCO, 1998). Furthermore, an increasing number of reports demonstrate that even those who have attended school for a few years may not have acquired and retained a minimum of basic skills such that they could be called "functionally literate".² This situation has implications for economic development of countries, as well as creates extra pressures on (and increase

costs of) health, employment and social welfare systems. However, in many developing countries there is no social or educational policy that aims specifically at the target groups of out-of-school youth and young adults.

Access to credible data about the status of basic learning competencies can offer policy makers and planners several advantages, and enable them to:

- Judge the current status of basic, functional and life skills within the out-of-school youth and young adult population, irrespective of former school attendance;
- Identify skills deficiencies among out-of-school youths and young adults (or subgroups within these populations) that have economic or societal implications and that can serve as targets for interventions;
- Know more about the relative effectiveness of existing formal and non-formal programs.³
- Make further progress towards meeting the 1990 Jomtien goal of reducing illiteracy by 50%.

Until the 1980s, most countries (rich or poor) have neglected the assessment of BLCs in their populations, partly due to a belief that universal education would guarantee a universally skilled population, and partly due to a lack of technical know-how in conducting BLC surveys. Instead, most countries relied on the use of surrogate or "proxy" measures, such as the assumption that any person with 5 (or 8) or more years of education can be considered literate. As has been often pointed out before, these types of measurements are relatively easy and cheap to obtain, yet are inaccurate and can be very misleading.⁴ Moreover, the scope of these early studies was limited to literacy, and rarely touched on broader issues of numeracy or life skills.

This report outlines a low-cost, culturally sensitive assessment framework that combine elements of household surveys (e.g., using moderately sized samples) with the use of measurement tools that are attuned to local and national needs. While maintaining a low level of operational and human resources costs, such assessment designs can satisfy the needs of international and national agencies for credible data as a pre-condition for supporting or investing in new human development initiatives. Further, these data can be also used to provide impact or evaluation data about national and local programs that teach basic skills (see section 6.1).

2.3 Challenges for policy-driven assessments

A number of challenges are involved in the monitoring of learning achievement and skill levels among the many millions of out-of-school youths and young adults who have had little or no formal education. Some of these are:

- In-school assessments (i.e., of students currently attending school) normally focus on the extent to which students have mastered some aspects of the common curriculum. In contrast, few curricula exist for people who are out of school, and there is lack of agreement as to what constitutes "basic" skills that are needed of individuals living in diverse contexts. Hence, it is unclear what should be the focus of assessments of out-of-school youths and adults. Further, out-of-school youth or young adult populations often cannot be assessed using school-type written assessments, due to limited formal skills, and assessments have to focus on realistic tasks or life contexts in which the target population has developed functional skills despite low formal skills.
- The results of in-school assessments can be useful if there is a way to directly link them to needed policy or implementation changes. However, many countries do not have explicit policies regarding out-of-school youth and young adults, or regarding education for life skills (especially beyond the primary grades). Hence, it may be unclear how to design assessments so that they have a chance to influence policy pertaining to the skills needed by young adults.
- Out-of-school youth and young adult populations often cannot be surveyed using the same sampling methodologies that are commonplace when studies of school-attending learners are conducted, such as by going to a school and assessing a classroom of students. By contrast, out-of-school youth and young adults have to be recruited or reached by different mechanisms, such as house to house surveys, or workplace surveys.

Lack of credible data about the breadth and depth of the "basic skills problem" has hampered the ability of policy makers to design and plan investments in programs or initiatives that can respond to the changing needs of out-of-school youth and young adults, especially those in disadvantaged circumstances. Having knowledge about the population distribution of various skills, competencies and attitudes among youth and young adults with different backgrounds and characteristics is a precondition for deciding how to promote literacy, numeracy, and important life skills in a way that is consistent with overall policy objectives.

2.4 What might policy makers wish to know about BLCs?

Policy makers are likely to commission a BLC survey in order to answer the following kinds of questions: (1) what does a given population 'know' in terms of BLCs; and (2) what policy changes can be made that might influence future levels of basic learning competencies or ways to improve the acquisition of BLCs. In addition to collecting skills-related data through the testing instruments included in a survey, collection of background information is essential in order to evaluate the determinants (independent variables) associated with observed skills (see section 4).

If policy makers wish to understand how different levels of prior schooling impact on actual (tested) knowledge or performance levels in certain skill domains, they need to make sure to have reliable data about relevant background variables, such as prior educational experiences (see section 6.1).

Policy goals will likely vary across countries, as well as by region and ethnic groups within countries. For example, countries concerned about overcoming inequalities caused by geographical disparities may want to collect more information about type of housing or community, or distance from nearest school, in order to be able to identify those living in remote or in inaccessible areas. Countries concerned about inequalities between different ethnic or language groups will need relevant variables that capture group membership in this regard. Finally, countries interested in relating the assessment results to specific programs or educational experiences respondents may have to include relevant questions concerning program curricula.

3. DEFINING BASIC LEARNING COMPETENCIES (BLCS)

Following the Jomtien perspective, BLCs should be discussed not only in terms of mastery of the 3 R's, but in terms of literacy and numeracy broadly conceived, as well as other knowledge and life skills and related dispositions. Together, BLCs are thought to promote empowerment and access to a rapidly changing world. They support independent functioning and coping with practical problems or choices as a parent or worker or citizen, and BLCs together are seen as critical gatekeeper to job entry and societal advancement in all countries. Thus, when defining BLCs, there is a need to refer both to formal school-based skills (such as ability to read prose text or to understand mathematical notations) and also to ability to manage functional tasks and demands, regardless of whether such competencies were developed through formal or nonformal education, or through learning from personal experiences in diverse informal settings.

3.1 Challenges in defining BLCs

Consensus has never been easy in agreeing on what 'people should know' or what skills define an 'educated' person. Over the years, a number of challenges consistently appear in the specialized literature in this area, such as:

- It is difficult to establish agreement among diverse, interested stakeholders who may hold different views of what constitutes the basic set of literacy, numeracy, or life skills in the context of a specific country or cultural group, and how to operationally define each skill and competence level. Consensus becomes even more difficult when definitions are expected to be valid and culturally relevant across two or more countries.
- The interest in performance in functional contexts often leads to defining multiple dimensions of literacy and numeracy, in relation to print, health, survival, workplace tasks and so forth. Yet, identifying functional literacy requirements or practices, and identifying life skills demands and needs, is not easy as these may not be stable over time or over geographical contexts within any given country.
- In many countries there exist multiple official and local languages and writing systems (or orthographies). Resource limitations often force decision-makers to conduct an assessment only in one or more dominant ("national") languages. This leads to situations where a person may be found to be "literate" or "numerate", or have effective life-skills, when assessed in one language, but the same person may be judged otherwise if assessed in a different language.

Consequently, conceptual definitions of BLCs can be posited, but cannot be expected to be the same across countries or cultural groups.

3.2 Direct and indirect measures

As noted above, definitions of BLCs are necessarily relativistic and culture-bound. Further, such definitions cannot be made in the abstract, but depend on decisions on the specific measures used to assess BLCs. In this regard, two key types of measures, direct and indirect, continue to be used in both testing and surveys of BLCs. Each is briefly described below, and each will be utilized in the proposed model for BLC assessment (described in detail in Section 5).

Direct measures of literacy and numeracy are based on the assumption that a person's ability can be best determined only if he or she is asked to actually perform certain specific tasks or practices. For example, direct testing of reading ability would require that a person is able to read certain words (which involves recognizing letters and decoding words), or that the person is able to read with understanding a simple sentence about

an everyday activity. More sophisticated forms of direct measurement techniques present respondents with a test or series of tasks, to capture different facets of reading, writing, numeracy and certain other life skills competencies. Tasks may be abstract or school-like, or may be functional in nature and use realistic texts (simulated or genuine newspaper articles, book excerpts, documents and official forms) to represent real-world stimuli as well as tasks found in people's home, shopping, community, or work situations.

Indirect measures typically estimate skill or knowledge levels based on self-report or surrogate (proxy) information, rather than on direct testing of skills. This is often done for methodological and cost expediency, relative to direct assessment. Presently, many countries report national literacy data via census questions such as: "Can you read and write?" In many countries, literacy status is also determined based on information about school attendance: those with at least five (or eight) years of formal public schooling are assumed to be "literate," and those with lesser levels are labeled as "illiterate." As noted elsewhere (ILI/UNESCO, 1998), it is now generally agreed that such approaches to indirect estimation of literacy levels can be quite misleading.⁵ On the other hand, for some BLCs, such as life skills, indirect measures based on self-report may be the most efficacious (or possibly only) place to begin assessment work, as many aspects of such skills often cannot be assessed as skill levels, since they encompass beliefs and attitudes as central components (see section 3.6).

3.3 Conceptual overview of assessment matrix

BLC skill assessment may be considered as consisting of various types of direct measurements that are determined by a matrix with two dimensions or axes: skill levels of BLCs (reading and writing, numeracy, and life skills) and skill domains (contexts or loci where the skills of literacy are applied). Four skill levels are defined, followed by domains.

Skill levels. To inform policy making, the measurement of literacy (reading and writing), numeracy, and life skills can be conceptualized as an attempt to place individuals on a continuum that is (for simplicity of description) comprised of four levels of ability: *none* or no ability in the designate area, *prerequisite* level, *basic* level, and *advanced* level.⁶ Each level is described below.

- *None*, inadequate or no ability level, refers to those individuals who, for all practical purposes do not possess even the rudiments of the skill domain in question. In the case of reading, for example, such persons could not write their name or even recognize letters of the alphabet.

- *Prerequisite level* refers to possession of specific knowledge or specific subskills that support performance in some set of functional learning tasks. Skills at this level provide the building blocks for future learning. However, prerequisite skills alone do not cohere into a functional, basic learning competency. In school-based terms, these isolated prerequisite skills are often part of the teaching content or curriculum of basic schooling, literacy campaigns, or informal programs.
- *Basic level* defines functional *competency* in performing common applied tasks that involve any of the BLCs, as well as potential for future skill growth. Specifically, this level should be thought of as reaching a stage at which prerequisite skills have been retained and can be put to functional use in a more integrated way, and upon which more advanced skills may be developed. This level represents a reasonable approximation of the spirit of *basic* notion suggested in Jomtien, though each country may need to define this level as best meets its own policy needs.
- *Advanced level* refers to a level of skill equivalent to that normally achieved upon successfully completing a secondary school curriculum. However, not all secondary school graduates actually meet this level, and some persons without such a diploma do manage to reach this level of attainment, based on other experiences they have had. A person who is at an advanced level should also be able to integrate separate skills and subskills in order to meet changing social and economic demands.

Within any level described above, one can of course define and measure multiple subskills. Hence, each country will need to decide on the skill range it should examine.⁷ For example, as elaborated further in section 5, reading may be divided into three measurable subskills, such as decoding, comprehension, applied skills. On the other hand, with respect to life skills, as discussed further below, it is clear that this new area of assessment cannot so easily be described in a common fashion across cultures, nor can the level of ability in life skills be as neatly put on a dimension of increasing competence. The present effort in producing a scale for life skills must be seen as a preliminary effort, and one that is closely tied to sectors such as health and nutrition

Skill domains. In addition to skill levels, there are multiple *domains* where skills are applied in everyday practice. For example, in reading, there are domains such as prose text (e.g., reading a simple relevant newspaper story, identifying key points in a political pamphlet) and "document" text (e.g., reading and making sense of official forms; directions on medicine bottle; advertisements; bills and receipts). These domains may be best understood within the major skill areas described below.

3.4 Defining literacy (reading and writing)

Many definitions exist for literacy. All relate in some way, at their core, to a person's ability to understand printed text and to communicate through print. Two well-known and oft-repeated ones are:

"A person is literate who can with understanding both read and write a short simple statement on his everyday life...A person is functionally literate who can engage in all those activities in which literacy is required for effective functioning of his group and community..." (UNESCO, 1978)

"Using printed and written information to function in society to achieve one's goals and to develop one's knowledge and potential. (OECD/Statistics Canada, 1995)

These definitions view literacy in relative rather than absolute terms. They assume that there is no single level of skill or knowledge that qualifies a person as "literate", but rather that there are multiple levels and perhaps types of literacy (or literacies). Definitions of literacy have to be sensitive to what constitutes literacy in *out-of-school* contexts, as well as to school-based skills. Thus, school-based skills such as grammar and punctuation may be of secondary importance in the everyday lives of many youth and young adults.⁸ In the present model we distinguish two measurable aspects of literacy - reading and writing - which, while surely related, may be observed and measured in different ways.

Some examples of the types of reading skills to be included in the present assessment model would include the following:

- Ability to name letters or smallest script unit (e.g., Chinese characters).
- Sight word recognition of common words in contextual print (road signs, package labels).
- Decoding skills - the knowledge and skills to apply decoding in recognizing and pronouncing written words.
- Comprehension of simple phrases and sentences.
- Ability to locate, comprehend and use information embedded in prose text or in documents.

Some examples of the types of writing skills to be included in the present assessment model would include the following:

- Ability to form letters and words using common writing implements (e.g., pencil and paper)
- Awareness of print conventions (directionality of script; top and bottom of documents; naming of punctuation.)
- Ability to copy a written text.
- Ability to write a text from dictation.
- Ability to compose a message, or to fill out a form.

In sum, any measurable definition of literacy will require distinct operational definitions for reading *and* writing skills.⁹ Thus, the assessment scheme for literacy, as described in section 5, elaborates separate tools for reading and writing. Furthermore, prerequisite and basic levels of literacy cannot be defined in the abstract, but rather the definition, within any cultural or national context, will be bounded by the characteristics of the language(s), script(s), and socio-cultural contexts of interest.

3.5 Defining numeracy

Most definitions of literacy have traditionally included calculating skills as part of the definition of literacy, but these have been limited primarily to the four arithmetic operations. By contrast, numeracy is viewed here as a separate and more comprehensive skill domain within the BLC assessment.

Numeracy encompasses a broad range of skills, thought processes, background knowledge (formal and /or informal), and certain attitudes and dispositions.¹⁰ Numeracy enables interpreting, acting upon and communicating about mathematical information in a wide range of everyday or work-related and other life contexts, and thus supports effective participation in a wide range of life roles.

Overall, numeracy enables critical comprehension of quantitative statements about trends and changes in one's community or nation, and in the world. Numeracy is needed as well for effective functioning in a world of amounts, volumes, shapes, drawings, maps, graphs, rates, prices, weights, distances, formulas, shipments, timetables, schedules, production statistics, and so forth. From a life skills perspective, numeracy is often required for the effective comprehension of information about health-related risks, side-effects of different treatments, and dosages. Finally, formal (school-taught) mathematical skills is often a prerequisite for entering diverse types of vocational or on-the-job training.

Some concrete examples of numeracy skills would be the following:

- Ability to count, sort, and perform simple comparisons of quantities of concrete objects (e.g., bottles, cattle, coins).
- Decoding the meaning of numerals (e.g., how many units of something are signified by the digit "5") and of number-words in an indigenous number system.
- Possess number-sense (awareness of meaning of numbers, and of magnitude of small and large numbers).
- Ability to carry out the four arithmetical computations, either mentally and/or in writing (using written symbols, notations, and procedures).
- Knowledge of measurement systems and procedures (telling time, using a scale, using a ruler or measuring tape, using money, converting between unit systems, reading dials).
- Understanding of percents, averages, graphs, and other means for summarizing and displaying data that are often used in conveying information about groups or about trends.

In sum, numeracy - in both formal and informal contexts - is increasingly understood to be a vital BLC which goes far beyond previous conceptions of school-based arithmetic. From personal empowerment to social and economic development, numeracy may be seen as a basic skill rivaling reading and writing in importance.

3.6 Defining life skills

The term "life skills" covers a broad cluster of competencies that have received little attention from educators as part of the formal school curriculum. In recent years, however, such skills are recognized as requiring much more consideration, in both developed and developing countries. In broad terms, the notion of life skills relates to the diverse knowledge, skills, values, and attitudes that together enable children, youth, and adults to prevent or cope with situations in which their own (physical or psychological) well-being or safety may be at risk, achieve their personal goals, function effectively in their social environment, and enhance their quality of life. According to UNICEF, "Life Skills enable children and young people to translate knowledge (what one knows) and attitudes/values (what one feels, believes) into action (what to do and how to do it)."¹¹

While there may be common agreement regarding some life skills, there is necessarily considerable diversity in specific subskills and knowledge components across societies. Some concrete examples of life skills might include the following:

- Knowledge of nutrition and eating habits (such e.g., is it safe to eat fish which is kept in a refrigerator for 3 days?).
- Knowledge of prevention of disease (e.g., What are effective ways to prevent getting HIV-AIDS).
- Knowledge of substance abuse (e.g., In what ways is tobacco smoking is dangerous to your health?).
- Knowledge of personal finances. (e.g., How much savings would be required to accumulate enough to build or buy a house in your village?)

Clearly, what constitutes "life skills" may differ depending on a person's age, place of residence, environmental and economic conditions, cultural practices, and other factors. Questions concerning knowledge about healthy eating will naturally vary depending on whether the person is a 10 year old youth or an adult, or whether he/she lives in an urban township or in a rural subsistence economy. Similarly, safe ways to cross streets, swim in the sea, or avoid snake bites depends on whether there are streets, beaches, or snakes in one's environment.¹²

Another challenge in assessment of life skills is in collecting useful and valid information not only about factual knowledge people may have, but also about attitudes, beliefs, and

values which have important links with actual healthy or safe behavior. Given that there is a relative dearth of research on assessment in the area of life skills, the assessment framework proposed for this BLC domain in section 5 focuses primarily on areas that we know how to measure (e.g. "knowledge about" questions in the examples above), rather than attitudes, beliefs and values which are so culturally dependent that a generalizable assessment plan would seem to be premature. In addition, we have little evidence as yet that life skills can be measured in terms of the four skill levels proposed for other BLCs, though we expect that future studies will provide a solid answer to this question. Thus, even though the present effort in measuring life skills should be viewed as ambitious and preliminary, the importance of life skills is widely acknowledged as central to the everyday survival, health, and well-being of millions of people.

4. SAMPLING AND SURVEY DESIGN

The challenges of designing a sampling scheme and survey instrument for any target population of out-of-school youth and young adults are formidable, yet worth the effort when considering the information gain over previous strictly census- or school-based surveys.¹³ This section begins with the proposed household survey design, and is followed by some case illustrations, and a number of principles for carrying out the survey successfully.

4.1 Proposed low-cost BLC household survey design

The proposed survey design described herein is derived from various previous examples which have combined background surveys with assessment instruments (e.g., UNSO, 1989; Greaney, et al., 1999; see also several case study illustrations, provided in Annex 3). While there are many aspects to the development of the instruments to be utilized, such as the principles described further on in this section, the three main components of the design may be seen in the survey instrument itself:

Assessment instruments for BLC skills. The direct assessment of BLCs consists of a series of items that will measure competencies in various domains. The development of these assessment instruments is a complex task, and must be carefully pilot tested to assure the usual degree of reliability and validity. Issues such as the type and number of test items, how they are to be analyzed, and so forth, are described in section 5.

Attitudes, values and dispositions questions. As described in section 5, the measurement of life skills will require a separate instrument that determines the individual's perspectives (in addition to skill levels) in various domains of human behavior (e.g., attitude toward having more children). While these questions may be coded in numerical form, they are not, in statistical terms, ordinal data that can easily be ranked in terms of what is 'best' or 'worst'.

Background variables. Background information on the demographics of the individual surveyed are critical for cross-tabulations to be made in the data analyses. For example, to make comparisons of BLC skill by gender across different regions, but while holding educational level constant, is not especially difficult if the requisite information is collected at the time of assessment.

Useful background items in household surveys may include at least the following types of variables:

- Personal data: age, gender, location/region, (e.g., urban-rural), major languages spoken, ethnic background
- Family data: size, educational status of parents, occupational status of parents, income, languages spoken at home, availability of literacy resources in house
- Education data: prior schooling, type of school attended, other educational/training experiences (whether in formal or non-formal contexts), language of instruction
- Employment data: employment status (yes/no/partial/seasonal), duration of employment/unemployment, wages
- Other: vocational or employment preferences, health status, health-relative habits or practices, etc.

As described elsewhere in this report, a considerable amount of preparatory work needs to be done in order to come up with a satisfactory survey design instrument that is ready to be administered across a variety of contexts in a given country. Each of the three component mentioned above are crucial. While most of the emphasis in this document has been placed on assessment issues, the background questionnaire (and its effective administration in the field) is crucial in terms of providing the right kinds of policy answers once skill assessment has been achieved. Overall, the advantages of this model of survey design are numerous, as described below.

4.2 Some advantages of this household survey design

The approach suggested in this document is that of a household survey design that is based on a population sample in a given country or region of a country. Embedded in the household survey is a BLC assessment at the individual level. There is ample basis for this choice, in terms of cost and efficacy, and numerous previous examples to follow upon, as illustrated in Annex 3. Nonetheless, it is important to enumerate why other approaches may be less suitable to BLC assessment of out-of-school youth and young adults. Three of the most important are as follows:

- Large scale surveys with international comparability, such as the IALS (1997), typically require greater cost in terms of sample size, expensive experts and infrastructure, and demand considerable time for development and analysis (see also footnote 18).
- Indirect measures of BLCs, such as those used in census questions about literacy, produce low reliability and validity, so that even an expensive full-sample population census may reveal little about the true state of literacy or other BLCs (see also section 3.2).
- School-based surveys clearly do not work well for an out-of-school population, even though some assessment items may be able to be 'transplanted' to contexts out of the school. School-based surveys are typically administered to a whole class at a single time, while out-of-school assessment usually requires a one-on-one approach to testing.

By contrast, the household survey approach to direct BLC assessment, used for example in Zimbabwe (United Nations, 1989) and in Bangladesh (Greaney, et al., 1999), provides the following positive elements, to summarize:

- Just enough information to answer key policy questions, thus a limit on the sample size needed.
- Direct measures of BLCs, leading to greater reliability and validity, even if more expensive for item development than indirect methods, and more expensive for one-on-one testing than in classroom testing (see cost section below).
- Greater focus on local culture and context, with less overhead costs in terms of international comparability.
- Greater data transparency for the non-expert user, and more "shareability" between national and program levels.
- Because most household surveys are individually administered and scored, it may be possible to include open-ended response items when the survey designers are uncertain that specific categories will capture all relevant responses.¹⁴

In sum, there are numerous reasons to choose a household survey design. But even within such a household survey, there are a variety of design and sampling considerations that need to be taken into account.

4.3 Principles for effective household survey designs

Difficult decisions often must be made regarding how many and how specific to make survey items. The following guidelines review

some key principles in designing surveys for assessing BLCs, including how best to undertake population sampling

Learn about the target population(s). Out-of-school youth and young adults should be able to respond directly to interviewer survey questions. Nonetheless, there are multiple factors that will influence (and potentially reduce) the reliability of responses, including language, culture and other factors.¹⁵

Define a target population, then select a representative sample of that group. Focus on achieving a representative sample of the targeted subgroups rather than population estimates of the entire national population. This strategy is in contrast to typical national population census sampling, which may or may not attain a representative sampling of specific subgroups. Furthermore, drawing representative samples of out-of-school youth may pose special sampling problems.¹⁶ A BLC assessment survey can also be built upon an on-going or upcoming national census (see ILI-UNESCO, 1998, for a description).

Small samples may be sufficient. A major cause of cost-inefficiency is selecting samples based solely on needed percentage of the overall population rather than on calculating the size of sample necessary to make reliable estimates of distributions in target sub-populations. One important principle of sampling that is often overlooked is that a properly selected randomized (yet small) sample can be much more representative of its parent (target) population compared to a larger sample selected by non-randomized methods.

Prepare planned comparisons. It is important to identify in advance for planned comparison groups or variables, such as by language/ethnicity, rural versus urban, economic status, and so forth. This approach will help determine which cultural, educational, home life, and other context variables to include in the survey instrument. Depending on the goals of the study, reliability of information can be increased by collecting secondary or corroborating data from parents or guardians, education agencies, employers or other groups who can confirm the evidence collected at the individual level. Continuous (ratio) data is typically more robust than interval, ordinal and nominal data. For example, date of birth is preferred over a range of ages (e.g. 20 to 25 years). Categories can be formed later, whereas international or longitudinal comparisons become impossible if category ranges or definitions change.¹⁷

Limit the scope of the survey. A manageable plan is required, with delimited sample sizes based on target groups. Planning for minority or disadvantaged group sampling, for example, entails numerous complexities. Over-sampling may be necessary to ensure adequate numbers among smaller and dispersed groups. Multilingual administrators, multi-language forms, travel and local customs must all be accommodated. Additional pilot testing of procedures before full implementation may be required. Ethnographic or case

study approaches may be a useful supplement when reliability of the survey poses a threat to survey generalization.

In sum, survey designs for BLCs purposes should focus on:

- Quality rather than quantity of information.
- Meaningful, interpretable data rather than broadly descriptive data.
- Particular populations (and comparison groups) rather than national representations.

4.4 Cost considerations

Each design consideration mentioned above has its associated costs. Throughout this document, we have referred to the present approach as one that is "low-cost." Clearly, the intent here is that such BLC assessment can be low-cost relative to other, "higher cost" approaches. It should be said at the outset that there are no absolute certainties with respect to cost, but there are some trade-offs that can be made in terms of cost and quality. The following are positive cost considerations with respect to the present type of BLC survey design.

- Limited sample household surveys can save money because they simply reduce the number of individuals needed to be interviewed in order to answer a set of particular policy questions. As noted earlier, larger scale studies, and international comparisons only can drive the costs upward.
- Lower levels of statistical reliability than sometimes used in censuses may be acceptable in BLC assessment in order to speed up data collection and reporting.
- Transparency and simplicity of the proposed survey design means that fewer experts are required to 'interpret' (or reanalyze) the data for a non-expert audience.
- Rather than being a 'one-off' (one time only) survey, the notion of 'shareability' (using the BLC design for both national and local program evaluation studies) can produce economies of scale. One-time-only surveys bear a considerable cost as well in terms of lost expertise the project team is disbanded after the project is complete, since complex surveys are typically impossible to replicate on the local program level.

5. MEASURING BLCS: AN ASSESSMENT MODEL

A wide variety of methods for assessing literacy and numeracy have been developed over time, with most focused on school-based skills. Less has been undertaken in a low-cost modality as described above, and very little in the way of methods have been developed in the Jomtien-fostered areas of life skills measures, oral expression and problem solving skills. The appropriate choice of assessment tools will depend not only the agencies involved, but also on the technical adequacy - or the empirical 'carrying capacity' - of the methodologies employed.

5.1 Levels of BLC skills

A key goal in BLC assessment is the development of levels of ability (or proficiency) scales that may be used to categorize BLCs in individuals and across groups. As noted earlier, many approaches have been tried to date - from the simple dichotomy of literate vs. illiterate, to continuous scales which can provide quite specific scores. The approach proposed here is a model, which draws on advanced statistical techniques, while maintaining a 'transparent' (i.e., clearly and simply understood) approach to test item selection and measurement. This model also assumes that countries (and education programs and agencies) will want to ascertain "basic" level targets of learning competency (i.e., what is the minimal goal of a "what should be learned or known" in basic education).

Thus, in the BLC skills described below, each is partitioned into four criterion based categories as described in section 3. The choice of four levels (rather than more or fewer levels) is related the need to keep cost down, while maintaining the possibility of reasonably delineated categories useful for policy and program development.¹⁸

Establishing test reliability can and should be undertaken in any model of BLC assessment, and this poses relatively few problems for specialists. Validity, on the other hand, requires not only the judgment of content experts as to the value or relevance of knowledge and skills tested (as related to policy goals), but also the test's ability to discriminate between individuals with varying levels of proficiency.¹⁹ Establishing internal consistency needs to be an iterative development process as countries seek to design measures of core competencies.²⁰ Many designers of school-based competency exams begin by gathering together experts and stakeholders (see section 6) to identify skill, knowledge, and value domains that constitute core competencies given an established curriculum. They then design a set of items that they hope can measure these competencies validly and reliably. However, the assumptions underlying the items must be tested. There are many reasons why a particular item may be easy or difficult for an individual other than his or her proficiency level in that skill. For example, the item can be ambiguously worded or culturally biased.

5.2 Reading Assessment

Literacy tests - the most common of BLC tests - have ranged traditionally from simple questions such as 'can you read and write', to signing one's name, to reading a short paragraph on a life-relevant topic, to answering multiple-choice questions on a test battery.

The BLC assessment scheme for reading is based on a matrix of reading skills and domains of print (see Table 1). This matrix, as described in section 3, can be used to define four ability levels: none, prerequisite, basic, and advanced. Reading skills, in this scheme are divided into three general categories: decoding, comprehension, applied skills.²¹ Three domains of print are described, including (1) prose text (e.g., newspapers, pamphlet, books, stories, etc.); (2) documents (e.g., official forms, labels, advertisements, bills, receipts, etc.);²² and (3) decontextualized print (e.g., letters, sublexical units such as syllables, words, phrases, and sentences).

None or non-reader level. This level refers to those individuals who, for all practical purposes, do not possess even the rudiments of reading skills, and cannot, for example, recognize more than a few letters of the alphabet at most.

Prerequisite level. Prerequisites to reading competency include letter recognition, decoding, and "sounding out" of short texts. Operational definitions of prerequisite levels of reading are somewhat difficult because of the variety of cognitive demands of learning the relationship of the written script to the oral language. For example, in some languages (e.g., Serbo-Croatian), there is a simple one-to-one correspondence between applying pronunciation rules to printed text and the spoken form of the language. In other languages, such as English or Arabic, the relation of printed text to oral language is not so simple and may require extensive knowledge of the linguistic, semantic, and grammatical structure of the language just to pronounce a printed text.²³ Thus, decoding skill must be operationalized with respect to specific language and script contexts.

Basic level. A basic level in reading ability can be defined as skill in "reading to learn" and "reading to do".²⁴ The former set of skills may be seen as most related school-based reading achievement, where the focus is on reading comprehension as a means for learning about content domains. The latter set of skills are more common to out-of-school functional literacy needs such as reading signs, following procedural directions, locating a specific item on a bus schedule, and other applied tasks.²⁵ Individuals with basic level competencies would be able to:

- Answer literal comprehension questions about brief, continuous texts (sentences and paragraph in length),
- Locate or apply information in simple but authentic prose pieces or documents, and
- Form judgments relating text information to background knowledge.

Advanced level. Advanced skills are built on those used in basic level tasks, but are applied to more complex tasks and print domains.²⁶ As noted earlier, advanced skills are equivalent to a

level of skill for those who have successfully completed secondary school curriculum or its equivalent.

5.3 Writing Assessment

The assessment scheme for writing is based on a matrix of writing skills and domains of print. This matrix, shown in Table 2, can be used to define four ability levels: none, prerequisite, basic, and advanced. The areas of skills are described below, followed by the four ability levels for writing assessment. Writing skills are divided into four categories: specific motor, script conventions, composition, and applied skills. Domains in writing, as with reading, would include both prose texts and documents. Culturally representative and familiar materials are essential for valid assessment of writing. For functional tasks, authenticity is important, even if it increases complexity.

None or non-writer level. This level refers to those individuals who, for all practical purposes, do not possess even the rudiments of writing skills, and cannot, for example, write more than a few letters of the alphabet at most.

Prerequisite level: Prerequisite to writing competency are specific motor skills, and knowledge of print conventions. Specific motor skills include the use of writing implements or "technologies" relevant to producing written language. Script conventions refer to knowledge of linguistic features as codified in the common use of written language (e.g., rudimentary punctuation, directionality of print, etc.). Since out-of-school youth may have had less experience using writing implements (e.g., pencil, chalk, pen, keyboard) and materials (e.g., paper, slate, chalkboard) such as those found in schools or many urban environments, actual writing skills assessed may be lower than the individual's potential if care is not taken to find and use writing technologies familiar to survey respondents.²⁷ Writing at this level can typically be measured by having the individual write to oral dictation, such as writing individual letters (or characters) of the alphabet or script. The next task would be to write common words that appear in their written form in functional contexts.²⁸ At a higher level, one may want to provide a graded list of words to write, varying the spelling complexity, as well as phrases and sentences.

Basic level. A basic level in functional writing ability can be defined as skill in "composing" and "applying" print (by hand or machine) such that a skilled reader could understand it.²⁹ The former (composing) set of skills is typically what is utilized and learned in school-based composition, where the focus is on writing as a means for communicating with others. The latter (applying) set of skills are more visible in out-of-school functional contexts, such as filling out forms, writing directions, making

lists, and so forth. Individuals with basic level competencies would be able to:

- Write a brief message to a literate peer;
- Write a brief description or story of an event or image;
- Fill out a simple form;
- Make a list of items from a familiar category, and
- Make a judgment of the functional quality (that is, interpretability) of a composed text.

Advanced level: Advanced skills would constitute the same as those used in basic level writing tasks, but are applied to more complex writing tasks. One might also apply a higher standard of print conventions such as command of punctuation, and sentence structure, as well as ability of complex writing activities (e.g., filling in a complex government form).

5.4 Numeracy assessment

The assessment of numeracy is based on a matrix of numeracy skills and domains, as depicted in Table 3. This matrix can be used to define four ability levels: none, prerequisite, basic, and advanced levels. Numeracy skills are divided into five categories: decoding, writing, computing, applying information, and measuring. Computing, locating information, and measuring are each necessary to basic learning competency in numeracy. One is not more developmentally advanced than another, though they have different contexts of use. Decoding and writing skills involve identifying, naming, or writing numerical signs, or symbols.

In contrast to literacy, it is essential that mental or oral (non-print) and informal numeracy tasks be part of the domains sampled for numeracy ability. In industrialized as well as in developing countries (or in cultures with no written orthography), people can count, make purchases, use money, measure quantities and distances, and so forth, even if they have no formal education or very low literacy skills. Further, the extent to which out-of-school youth/adults "know" formal math (i.e., what is learned at school) may not indicate very much about their overall numeracy competency, since they may have developed informal ways to handle specific mathematical tasks.³⁰ Thus, it is necessary to create more varied categories for numeracy than has been the case in most prior assessments which are largely focused on school-based math skills.³¹ Four levels are proposed to describe numeracy competencies (see Table 3c).

None or non-numerate level. This level refers to those individuals who have relatively few mental calculation skills beyond counting of simple quantities and who also cannot recognize the meaning of written digits.

Prerequisite level. The prerequisite level would include individuals who can engage in some (possibly even advanced) mental calculations using indigenous number systems or measurement devices/techniques only. However, they know few print-based or formal numeracy symbols and systems, though they may be able to do very simply written math problems.

Basic level. The basic level would include mental and written numeracy skill, including mathematical operations typically learned in school, as well as everyday math (such as estimates and graphs) as might be employed in newspapers. The basic level would thus be calibrated in terms of entry into employment training, economic, social and health-related needs. Perhaps even more than reading or writing, a basic level in numeracy may vary widely in different societies, depending on level of industrialization and urbanization, on the formal skill demands of basic education programs among out-of-school youth and young adults, or on other country-specific factors.³²

Advanced level. A person could be said to be highly numerate who can, with little difficulty or few errors, accomplish numeracy tasks as listed in Table 3c, such as perform various arithmetic functions in formal school-type and everyday settings, locate and use quantitative information in different displays, be able to solve multi-step calculations, use measuring devices, and so forth.

5.5 Life skills assessment

The assessment of life skills is based on a matrix of skills and domains of both knowledge and use (see Table 4). This matrix can be used to define four ability levels: inadequate, prerequisite, basic, and advanced levels. Life skills are divided (for heuristic purposes) into four categories: (1) knowledge, (2) attitudes/values/dispositions (AVD)³³ and actions, (3) critical awareness, and (4) problem solving. Knowledge involves the functional understanding and comprehension of content in a given life skill domain (e.g., smoking increases risk of lung cancer; how to read time on a clock). AVDs may be communicated and precede knowledge such as in the case of imposed parental or cultural rules or taboos (e.g., don't smoke, say 'no' to drugs), as well as "actions" previously taken that the individual acknowledges. Along with a sense of life "risks," critical awareness (being aware of risk) and problem solving (knowledge/skill about being able to 'handle' risky situations) are essential to basic learning competency in life skills. One caveat here is that, even more than in the domains of reading, writing, numeracy, the domain of life skills could be parsed in any number of ways. Our purpose here is to provide one practical schema for the development of assessment items, even knowing that others may be quite plausible.

Domains of life skills are listed in Table 4a. Unlike numeracy and literacy, there is presumed to be rather little in the way of transfer of skills across sub-domains of life skills. Knowledge, values, action, and problem-solving elements of life-skills that are self-reported or demonstrated in one context (e.g., health and health-related behavior, safety) imply relatively little about these same skills in another context (e.g., nutrition, substance abuse). Correlations among variables in this BLC area may be due to secondary variables, such as socio-economic status and education or cognitive capacity, rather than to a generalized "high" level of life skills. One implication is that valid measurement will require the use of a range of items that cover all the sub-domains and levels in each domain.³⁴

Inadequate (or high risk) level. This level refers to those individuals who, for all practical purposes, do not possess effective mastery of knowledge, skills, or attitudes as a basis for adequate life survival. In other words, such individuals (and their dependents) are at high risk to become involve in life threatening situations (e.g., HIV-AIDS, or other diseases and so forth), because they do not possess adequate understanding of life risks.

Prerequisite level: Prerequisites in life skills would include the knowledge, skills, and AVDs that are shown in Table 4a, but only in localized or constrained contexts. Knowledge and skills are information deemed fundamental to given a life skills domain. For example, in the domain of nutrition and safety, an individual may be able to identify specific safety signs in the road, but be unable to generalize to a broader array of high-risk signs and symbols in his/her environment.³⁵

Basic level: A basic level in life skills can be broadly defined as competency in acting and coping with critical life-risk domains, whether in health, everyday activities, and so on. For example, individuals with basic level skills in health and safety would be able to:

- Know/understand safety signs and rules, warning symbols, nutritional food groups;
- Demonstrate attitudes and values to not engage in risky health and safety behaviors;
- Demonstrate avoidance of risky behaviors, such as recognizing poor nutritional items in an infant's diet;
- Acquire information on or communicate a critical awareness of how and why such behavior can harm themselves, individuals, and society (e.g., attending a parenting class; asking advice; awareness high cholesterol foods and relation to health).

Advanced level: Advanced skills are the same as those used in basic level tasks, but applied and integrated into more complex tasks.

It is possible to define certain cognitive components (knowledge, skills) related to many life skills. However, it is

less possible to assess action based on life skills knowledge as part of a household survey. Several ways exist, nonetheless, to indirectly assess life skills:

- Ask respondents what is right (or proper) to do in a specific situation (e.g., being in a place where drugs are being offered, facing an aggressive act by a schoolmate);
- Ask respondents if they consider themselves able or confident to do a specific behavior; and
- Ask respondents how they would actually behave in a given situation.

Finally, problems of culturally appropriate content pose serious challenges to the valid assessment of life skills, given that no generic frameworks or item pools can be imported without major revisions. For this reason, there is a need for multiple item types for each life skill of interest. Keeping in mind that these items may not always be highly inter-correlated, there is a need to plan assessments so that there will be several items on each topic, examining not only knowledge and skill, but AVDs of the skill (such as the respondent's beliefs about what he/she should do, could do, and would do, regarding any given situation or topic).³⁶

In sum, while many policy makers and agencies are keenly interested in the measurement of life skills in youth and young adult populations, appropriate surveys have as yet been unavailable.³⁷

6. IMPLEMENTATION AND CAPACITY BUILDING

Implementation and capacity building are among the most familiar aspects of educational development. Yet, when applied to the terrain of BLC assessment, certain specific challenges arise, and options must be considered.

6.1 Design decisions and options for BLC surveys

Implementation of a household survey of out-of-school youth and young adults is not a simple undertaking and requires careful planning and often some capacity building. As discussed in detail in previous sections of this report, many decisions have to be made by policy makers and survey planners, in order to assure that:

- Questions of most interest to policy makers regarding BLCs of disadvantaged out-of-school youth and young adults are explicated and prioritized;
- Sampling methodology chosen for the survey can yield representative samples of the target population while keeping sample size to the minimum needed;
- Assessment instruments are designed to adequately cover the subskills, levels, and domains of literacy, numeracy, and life

skills in such a way that reliability and validity of the information obtained are sufficiently high;

- Survey is planned and implemented in the field in such a way that the information collected is credible and free of errors, while costs, logistical efforts, and time needed are kept at reasonable levels; and
- Design of data collection and reports from the survey are planned ahead of time to focus on variables and questions of maximum interest to the stakeholders and users of the results.

Planning processes need to be thought of as integral to the entire BLC survey. This is particularly the case if the data to be collected will serve both the information needs of national policy makers as well as decision-makers and directors of regional or local programs whose projects may serve out-of-school populations.

In contrast to national studies where policy makers are the main stakeholders, program level stakeholders would include directors and teachers in programs, and possibly employers, parents, and learners as well (who have a vested interest in the quality of the program they attend, and what they might learn from instruction). The early engagement and involvement of this diverse set of stakeholders needs to be discussed and determined in order to make sure a survey can answer as many important questions as possible that are raised by all such stakeholders, and in order to increase the chance that the results are considered credible and acted upon.

As described elsewhere (ILI/UNESCO, 1998), the notion of "shareability" of assessment tools is relevant here. Tools developed for national household surveys can and should be "shared" with program level evaluation and vice versa. Efficiencies can be gained in human resources if the same or similar assessment tools can be used to implement both national level surveys and local program evaluations.³⁸

Numerous challenges are involved in making sure that interviewing, coding, scoring, and analysis of responses can be conducted in efficient yet reliable ways. Hence, various phases of training (of interviewers, coders, data-entry operators) and quality assurance are necessary to ensure the accuracy and validity of the data.

6.2 Human resources: Recruiting and training personnel

The implementation of any complex survey of this kind depends on the availability of trained specialists, and on the building of the capacity of current personnel to undertake new tasks that require higher levels of expertise. One route to building human capacity is identifying stakeholder and client groups, engaging them in identifying their needs, and enlisting them to find qualified personnel. Sometimes sufficiently qualified personnel

can be found in the office (such as a census bureau) which undertakes the survey. However, this is rarely the case for BLC surveys, because census personnel often have little familiarity with psychometric testing or case study (qualitative) approaches to data collection. Similarly, a ministry of education office familiar with testing procedures may lack skills in collecting stratified samples of household level survey data. Furthermore, neither of these groups may have sufficient familiarity with the needs and demands for data on literacy programs and how reading skills (for example) relate to other life or numeracy skills. At the level of the survey or census team of field data collectors, the same must be said, as it is quite often at this level that the quality of the data is most crucial and least observable from the main office. Hence, the personnel required for a BLC survey usually necessitates a collaboration across offices, administrative boundaries, and types of survey specialists and workers.

Prior experience shows that additional specialists should be sought first, wherever possible, within national boundaries - primarily so as to build up national capacities.³⁹ When such human resources are not available within established ministerial agencies, such agencies should consider engaging national specialists who work in universities, research institutes or in the private sector. Multi-national collaborations should also be considered, especially when specific aspects of implementation call for experience from other countries.⁴⁰

6.3 Test development, pretesting and pilot studies

A "testing the test" development process is essential to enhance the reliability, validity, utility, and feasibility of the measures themselves before broad assessment is undertaken. Pre-testing and pilot-testing is needed to help identify numerous problem areas that could reduce information value and drive up costs.

To design cost-efficient measurement methods and instruments that have a high information value for BLC assessments, a number of guidelines are suggested, including:

- Develop standards of performance, measurement scales, and basic performance levels.
- Use functional, performance-based rather than academic, school-based tasks and tests.⁴¹
- Strive for contextual relevance rather than national comparability.
- Develop a specific set of task items that can efficiently measure BLCs, while acknowledging the fact that there is no universal test that has the same meaning for every subgroup.

- Collect measures of attitudes, values, and dispositions through self-report surveys rather than through proficiency tasks.
- Plan for an on-going process of test construction, revising the test, and validation, since the BLC measures (or revisions of them) will likely be used both at national and local levels over numerous years.
- Plan to share instruments and report analyses of strengths and weaknesses of tests and tasks across different kinds of users (such as census bureaus, NGO programs, national and international agencies).

Pretesting and pilot studies are cost-efficient ways to increase quality in all facets of BLC survey work.⁴² Assessment tools that may be reliable and valid in one context may require adjustment before being validly applied to a new context, cultural group or population.⁴³

Pilot studies and pretesting can also be used to reduce the number of survey items or tests necessary to achieve the reliability necessary to achieve survey goals. Since many BLC assessment items tend to be inter-correlated (e.g., reading and writing sentences), researchers often find considerable analytic redundancy.⁴⁴ Test items may also fail to discriminate sufficiently, are too easy or difficult for a given population, or have cultural biases that warrant revision or exclusion. Pretesting can help to cull these extraneous items from the pool of test items or point to items that need revision, reducing time and cost in the administration of the survey and reducing difficulties in achieving reliable scoring of responses later on.

Another important form of bias is related to whether the test itself may be seen by the population sampled as either 'high' or 'low' stakes. By definition, high stakes assessment is one in which results are used as a gatekeeper. For example, university entrance exams are an example of a high stakes individual assessment. Low stakes assessment, such as those presumably used in the BLC household survey, typically are not tied to individual advancement (since no diploma is being issued in the survey). However, it should be noted that in many developing countries, individual assessment of any kind may be perceived as high stakes, due to government involvement. One issue that is important in this context is whether to provide monetary (or other) incentives to survey participants.⁴⁵

In general, a household survey involves sending trained interviewers into people's homes to conduct one-on-one (or face-to-face) interviewing. Testing (and individual survey questions) should be undertaken in as quiet and controlled setting as possible. Responses should be written down by the interviewer verbatim for later scoring using accepted methodologies. When mother tongue language and the language of the test (or test administrator) are different, further checks on validity may be required. When testing is in a written script that is a second

language for the respondent, an oral translation should also be prepared in the first language wherever possible. After interviewing is finished, responses to test questions and to any background questions (e.g., about current employment and wages) have to be coded or scored and then entered into computer files to be analyzed.

6.4 Ensuring the quality of data collection

The quality of data collection is essential to the credibility of any survey, no matter how much care was taken in survey design, development, planning, and analysis. Before the study starts, several key topics have to be considered and appropriate solutions planned:

- Recruiting, training, and monitoring of field interviewers and of coders, scorers, and data-entry operators have to be designed and planned;
- Scoring forms and data inputting and screening processes for computer processing during the data collection phase must be planned in advance; and
- In diverse and multilingual contexts, established and cross-translated guidelines are required for acceptable and unacceptable variations in standardized administration of interviews (for example, checks of inter-rater agreement, in which different scorers give the same score to a certain response).

Once a study is underway, regular discussion and feedback between administration planners, design personnel, and data collectors are essential to ensure both the reliability and validity of the data. Even if rigorous pilot testing was conducted, some data processing should take place at an early stage of the data collection phase. Waiting until all the data are collected before checking for problems is risky, as problems that were not encountered during the pilot stage cannot be discovered until it is too late to correct them. Control data sheets should include cells for participants reached (and/or rejected) based on sampling design, participant survey forms key-coded to control forms, database structures, and so forth. A percentage of these forms should be cross-checked on a subsample of data before all data collection is completed.⁴⁶

6.5 Data processing and analysis

After raw data are entered in computer files, data screening (or checking) is also important and should be built into the implementation plan. Screening processes include searching for out-of-range values (e.g., people born on February 30), implausible means and standard deviations, missing data,

abnormality of distributions, data outliers, as well as multi-collinearity of variables.

At least three levels of data analysis should be utilized: (1) to establish reliability and validity of the survey items and instruments (the culmination of the instrument construction process); (2) to generate descriptive statistics that provide an overview of BLC skill levels, sub-indexed by major background and demographic variables of the survey (e.g. age, language, region); and (3) to address policy questions as well as forecasting, based on issues developed in the policy planning phase.

Multivariate statistics should be used wherever the data permit an analysis of competing independent variables (such as education and income). These analyses can become the basis for developing parameters for policy recommendations at the national or program level.

6.6 Reporting and dissemination

As noted earlier in this section, policy impact requires an understanding of the consumers (users) of reports on the BLC survey. Governmental agencies and program level stakeholders will likely have somewhat different interests in individual and subgroup profiles, as well as in the policy implications of data analyses and interpretations. Thus, it is likely that more than one report (or at least various sub-reports) will be required in order to address diverse constituencies.

Thus, even during the initial planning process of the survey, a strategic plan for using reports and preparing for dissemination of conclusions should be put into place. Wherever possible, these reports should be tailored to the goals and purposes of the consumer of these reports, with a special effort at showing how the results of population sampling and direct assessment provide a more credible and cost-effective method for policy planning in the field of basic education.

7. CONCLUSIONS AND NEXT STEPS

The present report describes a conceptual and analytic framework, along with methodological guidelines for the low-cost assessment of basic learning competencies (BLCs), especially designed for disadvantaged out-of-school youth and young adults. Its purpose is two-fold: (1) to serve as an up-to-date synthesis of household survey approaches to BLC measurement and assessment; and (2) to provide practical guidelines to what is required to accomplish such a survey.⁴⁷

The overall goals and purposes of BLC assessment should include the need to:

- Obtain better literacy statistics for international and bilateral agencies;
- Provide national decision makers with better data for policy making;
- Provide program directors and NGOs with improved ways of showing effectiveness on the ground; and
- Enhance capacity building to achieve the above goals.

Following on these goals, the present report suggests a coherent approach to the definition and assessment of reading, writing, numeracy and life skills, along with a design strategy for the implementation of a low-cost household survey approach. Various specific guidelines and suggestions about implementation are provided both in the text and in further examples in the footnotes, tables, and Annexes.

Based on this approach, a number of next steps are needed as follow up activities:

- Initiate a series of national pilot studies on BLC assessment. These studies could take the form of census supplement (improving the current census data on literacy), a calibration study (adding a small scale study to the census), and/or a full-scale household survey as described in this report. At a program level, which ideally would take place along with the national studies, there would be case studies and (if funding available) longitudinal studies designed to support work at the grassroots level.
- Within the context of these pilot studies, develop BLC testing instruments in the appropriate languages and with culturally relevant information. These methods would follow on the sample design and instruments described in this report.
- Efforts should be made to link education and other sectors (e.g., health, nutrition, agriculture, work, etc.) with the measurement of BLCs, especially in NFE programs.
- Capacity building must be central to BLC assessment efforts, since the range of skills required often exceeds that of a single government agency. Where possible, training (such as national or regional workshops) should be tied to specific concrete projects, in order to benefit from hands on experiences.

This report, conducted in light of the 1990 Jomtien Conference and in anticipation of the EFA2000 evaluation activities, is a reflection of the increased concern about the status of basic learning competencies in many nations, as well as about the effectiveness of education programs (formal and non-formal) at all levels. Through the use of BLC assessment models, such as those described herein, information collection for policy formulation and planning can be much improved.

TABLES

Table 1a: Reading Skills and Domains for Direct Measurement

Reading skills for direct measurement

Decoding (e.g., recognizing/pronouncing letters, sublexical units, familiar words)

Comprehension- "reading to learn" (e.g., answer literal understanding of phrase, sentence, paragraph questions; recall information; relate text to background knowledge)

Applied - "reading to do" (e.g, locating information in prose or documents; following directions or procedures)

Reading domains for direct measurement

Letters, words, phrases, sentences (e.g., labels of objects; signs)

Prose (e.g., newspaper story, written passage)

Documents (e.g., official forms; directions on medicine bottle; advertisements; bills and receipts)

Table 1b: Operationalization Matrix for Reading Assessment

	TYPE OF SKILL		
	<u>Decoding</u>	<u>Comprehension</u>	<u>Applied</u>
DOMAINS			
Letters, etc.	a	-	-
Prose	-	b	c
Document	-	d	e

Table 1c: Reading Levels for Assessment

i. Non-reader level

Cannot effectively use skills in Level ii.

ii. Prerequisite level, with some errors,

Can decode (identify/say) letters, words in isolation [a]*

Can decode (identify/say) common words in everyday context [a]

Can read aloud connected text (phrases, sentences, or simple, familiar passages) with appropriate pauses and intonation [a]

iii. Basic level

Possesses skills in Level ii, and, with some errors,

Can answer literal comprehension questions in a short text [b, d]

Can locate specific information in simple, familiar document [e]

Can complete a simple life skill task using print [c, e]

Can follow simple written directions [c, e]

iv. Advanced level

Possesses skills in Level iii, and, with some errors,

Can understand and critically respond to print information [b, d]

Can learn new information from a simple prose document [b]

Can locate information in a more complex, or novel document [d]

*Appropriate cells are listed in brackets []

Table 2a: Writing Skills and Domains for Direct Measurement

Writing Skills for direct measurement

Specific motor (e.g., writing letters, copying familiar words)

Script conventions and fluency (e.g., copying dictated sentences using some common print conventions; writing a grammatically understandable sentence noun, verb in order)

Composition - (e.g, writing a description or story to a visual or verbal prompt; writing a message to a friend)

Applied - (e.g, filling out a form; creating a list of items; making table)

Writing domains for direct measurement

Letters, words, phrases, sentences (e.g., label; sign name; phrase; sentence)

Prose (e.g., story, message, letter)

Documents (e.g., filling out a form; list; table)

Table 2b: Operationalization Matrix for Writing Assessment

	TYPE OF SKILL DOMAINS			
	<u>Specific</u> <u>motor</u>	<u>Conventions</u>	<u>Composition</u>	<u>Applied</u>
Letters, etc.	a	b	-	-
Prose	-	c	d	-
Document	-	e	f	g

Table 2c: Writing Levels for Assessment

i. Non-writer level

Cannot effectively demonstrate mastery of any skills in Level ii.

ii. Prerequisite level, with some errors,

Can copy with accuracy letters, words, short phrases or sentences [a]*

Can copy from dictation letters, words, short phrases or sentences using some print conventions of punctuation, capitalization, etc. [b]*

Can sign name [a]

iii. Basic level

Possesses skills in Level ii, and, with some errors,

Can write a simple query or message [d]

Can write simple directions [d]

Can write a list of items [f]

Can complete a simple, familiar form [g]

iv. Advanced level

Possesses skills in Level iii, and, with some errors,

Can compose a paragraph or more of text showing command of the conventions of print grammar and structure. [d]

Can fill out an unfamiliar, and moderately complex form. [g]

*Appropriate cells are listed in brackets []

Table 3a: Numeracy Skills and Domains for Direct Measurement

Numeracy Skills for Direct Measurement

Decoding. (e.g., recognizing and pronouncing written numbers)

Writing. (e.g., writing numbers from dictation; copying written numbers, filling forms)

Computing. (e.g., ability to perform the four arithmetic functions: addition, subtraction, multiplication, division; with numbers or with concrete objects or money; mentally or by written means)

Locating information. (e.g., finding the required numbers in a paragraph or document)

Measuring (e.g., telling time, using a scale, a ruler, reading dials)

Numeracy Domains for Direct Measurement

Written numbers. (e.g., single numbers; isolated numbers on signs)

Numbers in oral text. (e.g., numbers in oral discourse)

Numbers in prose text. (e.g., in a newspaper article, a political pamphlet, a medical brochure)

Numbers in documents. (e.g., in official forms; directions on a food can; advertisements, instruction manuals for machinery) Measurements (time, weight, length and distance, volume, etc.)

Table 3b: Operationalization Matrix for Numeracy Assessment

	TYPE OF SKILL DOMAIN				
	Decoding	Writing	Computing	Locating	Measuring, Graph
Written numbers	a	b	c	d	e
Numbers in oral text	--	--	e	--	--
Numbers in prose text	--	--	f	g	h
Numbers in documents	--	i	j	k	--
Everyday objects & actions	--	--	--	--	l

Table 3c: Numeracy Levels for Assessment

i. Non-numerate level

Cannot effectively demonstrate mastery of any skills in Level ii.

ii. Prerequisite level, with some errors,

Can decode (identify/say) numbers in isolation [a]*

Can decode (identify/say) numbers on signs in everyday contexts [a]

Can write numbers, copy written numbers [b]

Can solve simple "oral" arithmetic problems mentally [e]

Can read weight information off a label [d]

iii. Basic level

Possesses skills in Level ii, and, with some errors,

Can locate information and fill in forms and documents [d]

Can solve simple written arithmetic problems on paper [f]

Can interpret numbers on medicine label [j]

Can measure length of item with ruler [l]

iv. Advanced level

Possesses skills in Level iii, and, with some errors,

Can locate complex numerical information (in document, prose [k]

Can decide on price of an item after discount [j]

Can problem-solve (e.g., calculate length of fence around a building, based on drawing, and using a scale) [l]

Can utilize formal number systems, percentages, estimations [j]

*Appropriate cells are listed in brackets []

Table 4a: Life Skills and Domains for Direct Measurement

Life skills for direct measurement

Functional knowledge (e.g., recognizes the poison symbol or road warning signs; knows that water conducts electricity)

Risk avoidance attitudes, beliefs, dispositions, actions (e.g., practices good eating habits or safe sex)

Critical awareness (e.g., aware of gender issues that influence opportunities for work or education.)

Problem solving (e.g., knows how to resist peer pressure)

Life domains for direct measurement

Health/hygiene/nutrition (e.g., disease prevention and treatment; cleaning and dietary habits)

Everyday life (e.g., accidents and safety, home life)

Social environment (e.g., civic sense, caring, negotiating)

Natural environment (e.g., protection and awareness of environment)

Workplace and commercial environment (e.g., functional office skills)

Table 4b: Operationalization Matrix for Life Skills Assessment

	TYPE OF SKILL			
	<u>Knowledge</u>	<u>Attitudes,</u> <u>Beliefs,</u> <u>dispositions,</u> <u>actions</u>	<u>Critical</u> <u>awareness</u>	<u>Problem Solving</u>
Health				
Everyday life				
Social environment				
Natural environment				
Workplace				

* Each Domain is independent of the others, therefore levels cannot be generalized across domains.

Table 4c: Life Skills Levels for Assessment

i. Inadequate (or high risk) level

Cannot effectively demonstrate mastery of knowledge, skills, or attitudes in Level ii.

ii. Prerequisites level, within a given domain,

Can show some examples of functional knowledge of high risk issues, behaviors, and concepts [first column]

Can demonstrate some attitudes, behaviors, and actions that reduce the risk of personal or social harm [second column]

iii. Basic level

Level ii, and, with some errors,

Can demonstrate many attitudes, values, behaviors, and actions that reduce the risk of personal or social harm [second column]

Can show critical awareness of self in relation to important personal and social life skill issues [third column]

Can show problem solving skills in coping with important personal and social life skill issues [third column]

iv. Advanced level

Possesses skills and knowledge in Level iii, and, with some errors,

Can show high level of critical awareness of self in relation to important personal and social life skill issues [third column]

Can show high level and integrated problem solving skills in coping with important personal and social life skill issues [third column]

REFERENCES

- Carraher, T. N., Carraher, D. W. & A. D. Schliemann. (1985). Mathematics in the streets and in schools. British Journal of Developmental Psychology, 21-2.
- Chall, J. S. Stages of reading development. New York: McGraw-Hill Book Company, 1983.
- Chinapah, V. (1997). Handbook on monitoring learning achievement: Towards capacity building. Paris: UNESCO.
- Furniss, E. & Baldo, M., Life Skills: An Evolving Concept (draft paper). New York: Unicef.
- Gal, I. (1997). Numeracy: Imperatives of a forgotten goal. In L. A. Steen (Ed.), Why numbers count: quantitative literacy for tomorrow's America (pp. 36-44). New York: The College Board.
- Greaney, V. Khandker, S. R. & Alam, M. (1999). Bangladesh: Assessing basic learning skills. Washington, DC/Dhaka: World Bank.
- ILI/UNESCO, 1998. Literacy Assessment For Out-Of-School Youth And Adults. (ILI/UNESCO Technical Report from Expert Seminar, Paris, June 1998.) Philadelphia: International Literacy Institute, University of Pennsylvania.
- Khan, N. U. (1999). Total literacy campaign of District Bikaner: External evaluation report. New Delhi: Department of Social Work, Jamia Millia Islamia.
- Kirsch, I., Jungeblut, A., Jenkins, L., & Kolstad, A. (1993). Adult literacy in America: A first look at the results of the National Adult Literacy Survey. Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- Kirsch, I. & Mosenthal, P. (1990). Exploring document literacy: Variables underlying the performance of young adults. Reading Research Quarterly, 25, 5-30.
- Mauritius Examinations Syndicate. (1996). A survey of 9 year old children in Mauritian Schools in Literacy, Numeracy, and Life Skills. Mauritius: Unesco/Unicef Project Report.
- McLeod, D. B. (1992). Research on affect in mathematics education: A reconceptualization. In D. A. Grouws (Ed.), Handbook of research on mathematics teaching and learning (pp. 575-596). New York: Macmillan.
- Murray, T. S. (1997). Proxy measurement of adult basic skills:

Lessons from Canada. In Tuijnman, A., Kirsch, I. & Wagner, D.A., (Eds.). Adult basic skills: Innovations in measurement and policy analysis. Cresskill, NJ: Hampton Press.

OECD/Statistics Canada (1995). Literacy, economy and society. Paris: OECD.

Oney, B., Peter, M., & Katz, L.. Phonological processing in printed word recognition: Effects of age and writing system. Scientific Studies of Reading, 1997.

Tuijnman, A., Kirsch, I. & Wagner, D.A., (Eds.). (1997). Adult basic skills: Innovations in measurement and policy analysis. Cresskill, NJ: Hampton Press.

UNESCO. (1990) Final Report on the World Conference on Education For All: Meeting basic learning needs, Jomtien, Thailand. Paris: UNESCO.

UNESCO. (1996). Education For All: Mid-decade report. Paris: UNESCO.

UNESCO (1998). World Education Report. Paris: UNESCO.

United Nations Statistical Office (UNSO). (principal authors Wagner, D.A., & Srivastava, A.B.L.) (1989). Measuring literacy through household surveys. (Doc. No. DP/UN/INT-88-X01/10E). New York: United Nations Statistical Office.

US Department of Education. (1997) Adult Literacy in OECD Countries-Technical Report on the First International Adult Literacy Survey. Washington, D.C.: National Center for Education Statistics.

Wagner, D. A. (1992). Literacy: Developing the future. Unesco Yearbook of Education, 1992, Vol. 43. Paris: Unesco.

Wagner, D. A. (1993). Literacy, culture and development: Becoming literate in Morocco. New York: Cambridge University Press.

Wagner, D. A. (1990). Literacy assessment in the Third World: An overview and proposed schema for survey use. Comparative Education Review, 33, 1, 112 - 138.

Wagner, D. A. (1999). Rationales, debates and new directions: An Introduction. In Wagner, D. A., Venezky, R.L., & Street, B.L., Literacy: An International Handbook. Boulder, CO: Westview Press.

Wagner, D. A., Venezky, R. L., & Street, B. V. (Eds.) (1999). Literacy: An International Handbook. Boulder, CO: Westview Press.

World Health Organization (WHO). (1999). Guidelines: Life skills education curricula for schools. Geneva/ICI-Canada: author.

ENDNOTES

EFA 2000 Assessment Surveys Report

**ASSESSING BASIC LEARNING COMPETENCIES
AMONG YOUTH AND YOUNG ADULTS
IN DEVELOPING COUNTRIES:**

**Analytic Survey Framework
and Implementation Guidelines**

ANNEXES 1-4

Final Report Based on
Expert Workshop for EFA2000
April 12-14 1999, ILI-UNESCO, Philadelphia

ANNEX 1:

**PARTICIPANTS AT BLC WORKSHOP
Philadelphia, April 12-14, 1999**

RESOURCE PERSONS

Meng Hongwei
Research Center of Educational Measurement and Monitoring
CNIER, Bei Sam Huan Zhong Lu 46
Beijing, 00088 China
Tel: (86-10) 6 202 90 92
Fax: (86-10) 6 202 90 92
menghw@bj.col.com.cn

Iddo Gal
University of Haifa
Department of Human Services
Haifa 31905 Israel
Tel: (972) 4 825 2245
Fax: (982) 4 824 6832
iddo@research.haifa.ac.il

R. Govinda
School & Nonformal Education Unit
NIEPA
17-B, Sri Aurobindo Marg
New Delhi 110016 India
Fax: (91) 11 685 3041
govinda@del2.vsnl.net.in

Vincent Greaney
World Bank
1818 H Street, NW
Room D9105
Washington, D.C. 20433
Tel: 202 458 1221
Fax: 202 522 1662
vgreaney@worldbank.org

R. A. Manrakhan (MES, Mauritius)
Mauritius Examinations Syndicate
Reduit, Mauritius
FAX: 230 4 47 675
mesexam@intnet.mu

Scott Murray
Statistics Canada
Culture, Tourism and the Centre for Education Statistics Canada
17-B R.H. Coats Building
Ottawa, Ontario K1A 0T6
Canada
Tel: 613 951 9035
Fax: 613 951 9040
scotmur@statcan.ca

UNESCO

Vinayagum Chinapah
Head ED/EFA/MIP
UNESCO
7, Place de Fontenoy
75007 Paris, FRANCE
Tel: (33) 01 45 68 1000
Fax: (33) 01 45 68 5629
v.chinapah@unesco.org

Unicef

Elaine Furniss
Unicef
3 United Nations Plaza
New York, NY 10017
Tel: 212 702 7283
Fax: 212 702 7149
efurniss@unicef.org

Aster Haregot
Unicef
3 United Nations Plaza
New York, NY 10017
Tel: 212 702 6622
Fax: 212 702 7149
aharegot@unicef.org

International Literacy Institute (ILI)

University of Pennsylvania
3910 Chestnut Street
Philadelphia, PA 19104-3111
USA
Tel: 215 898 2100
Fax: 215 898 9804
ili@literacy.upenn.edu

Serafin Coronel-Molina
Coronel@literacy.upenn.edu

Christopher Hopey
Hopey@literacy.upenn.edu

Kelly Limeul
Limeul@literacy.upenn.edu

Mohamed Maamouri
Maamouri@literacy.upenn.edu

John Sabatini
Sabatini@literacy.upenn.edu

Leigh Swigart
Swigart@literacy.upenn.edu

Dan Wagner
Wagner@literacy.upenn.edu

ANNEX 2: SAMPLE ITEMS FOR ASSESSMENT MODEL

SUB-ANNEX 2.1: Reading Assessment Samples

Sample A - Prerequisite level reading tasks

	TYPE OF SKILL		
	DECODING	COMPREHENSION	APPLIED
DOMAINS			
Letters, etc.	[a]*	-	-
Prose	-	b	c
Document		d	e

Decoding x Letters, etc. (Adapted from WRAT - Jastak, 1993)

*Brackets [] indicate targeted skill in the matrix

Instructions (read aloud by administrator):

Look at the letters on this line (point). Read me the letters one by one.

Look at each of these words carefully. Read each word across the page so I can hear you. When you finish a line, go to the next line.

Stimulus (given to respondent)

A B R T C	M O H P E	F L U Y V	
at	fish	key	pen
what	pencil	finger	circle
below	guide	answer	newspaper

Scoring

Continue until 10 in a row are pronounced incorrectly.

[Note: the WRAT has developed a norm-referenced score key for judging levels. Country-specific mastery levels can be determined by a combination of content experts and pilot testing against comparison groups of known ability levels.]

Comments and Guidelines

Some test development guidelines for pre-requisite skills:

- order letters by frequency of use in language
- select words that appear in print frequently
- select words that are familiar oral vocabulary terms to most native language speakers

- vary parts of speech (verbs, nouns, prepositions) and use most common usage (present tense, non-plural)
- order words in increasing complexity by features such as length, number of syllables, decodability, abstractness
- pictures or line drawings can be used
- confirm that respondent knows meaning of word pronounced (in transparent orthographies), or when oral pronunciation may underestimate recognition (as in second language readers)

Alternate Item Formats

Say:

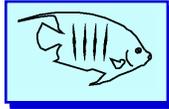
- Show me (circle) the letter R.
- Show me (circle) the letter M.

Answer Sheet:

- A B R T C
- M O H P E

Say:

2. Show me (circle) the word fish.
3. Show me (circle) the word that tells what is in this picture.
(include drawing of fish)



Answer Sheet:

3. fish key fast frog tree
-

Sample B - Basic level reading tasks: prose

	TYPE OF SKILL		
	DECODING	COMPREHENSION	APPLIED
DOMAINS			
Letters, etc.	a	-	-
Prose	-	[b]	[c]
Document		d	e

Prose Comprehension (Greaney, 1999)

Instructions (read aloud by administrator):

Now I am going to show you some pages, and you are to read what is on each page. Please read aloud or silently. When you are through reading a page, look at me or say "Done" to let me know that you are finished. I will then ask you a question about what you have read... Go ahead and read this.

Stimulus (given to respondent)

Father said that he would 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!

Instructions (questions read aloud by administrator):

1. Name two vegetables mentioned here.
Correct response: Any two of cauliflower, potatoes, tomatoes, or winter vegetables
2. How many days worth of food will they get from the land?
Correct response: Accept one year or equivalent
3. Whose smile is a source of pleasure?
Correct response: mother
4. What does that smile resemble?
Correct response: accept the grown harvest or the harvest

Scoring

Note: Set a mastery level of total to answer correctly to pass this level, then empirically validate with internal consistency methods.

Alternate Item Formats

You can assess fluency with this item type but it requires additional administrator training for inter-rater reliability.

Administrator instructions: Ask respondent to read aloud. Circle all words obviously mispronounced or skipped. Then rate reading fluency based on three levels.

- Dysfluent (many words are sounded out separately with long pauses and more than 5 errors)
- Understandable, with difficulty (less than five errors, no long pauses, but monotone word by word reading)

- Fluent (good tone, stress, and stops based on recognition of punctuation and language features.)

Comments and Guidelines

Some test development guidelines for comprehension tasks:

- write or select a passage that describes a content and context of high cultural familiarity
 - check that passage words are familiar, oral vocabulary terms to most native language speakers
 - order passages from short (sentence length), easy and literal to longer (paragraphs), and more abstract in content
 - pictures or line drawings may be used if they are part of an authentic prose or document piece. Otherwise visuals tend to confound reading with other visual literacy and interpretation skills.
-

Sample C-1 Basic level reading tasks: document

	TYPE OF SKILL		
	DECODING	COMPREHENSION	APPLIED
DOMAINS			
Letters, etc.	a	-	-
Prose	-	b	c
Document		[d]	[e]

(Adapted from Mauritius Exam Syndicate, 1997)

Instructions (read aloud by administrator):

Now I am going to show you a poster announcement. Please read aloud or silently. When you are through reading, look at me or say "Done" to let me know that you are finished. I will then ask you questions about what you have read... Go ahead and read this.

Stimulus (given to respondent)

<p>SOCCER COMPETITION AT THE NATIONAL SPORTS STADIUM ON FRIDAY 22 AUGUST FROM 12:00 - 17:30</p> <p>MANY PRIZES TO WIN!</p> <p>Come and enjoy the day Ice Cream, Soft Drinks, Fruits, Pop Corn will be sold</p>

Instructions (read aloud by administrator):

1. What is being announced in this poster?
 Correct response: Accept sports competition, or soccer match.
2. Tell me (Show me/circle) the start time.
 Correct response: 12:00
3. At the event, name two foods people will be able to buy?
 Correct response: Any two of Ice Cream, Soft Drinks, Fruits, Pop Corn
4. Where will the competition take place?
 Correct response: Accept the National Sport Stadium or a sports stadium.

Scoring

Note: Set a mastery level of total to answer correctly to pass this level, then empirically validate with internal consistency methods.

Comments and Guidelines

Some test development guidelines for document tasks:

- write or select a documents that describes a content and context of high cultural familiarity
 - check that documents use familiar, oral vocabulary terms to most native language speakers
 - order documents based on complexity of information from simple (e.g., a two column table) to more complex (e.g., a bus schedule with multiple columns and footnotes for interpreting special symbols.)
 - make tasks as realistic to the typical purposes and uses of the document as possible (e.g., look up the time a bus leaves; determine dosage on a medicine label.)
-

Sample C-2 Basic level reading tasks: document

	TYPE OF SKILL		
	DECODING	COMPREHENSION	APPLIED
DOMAINS			
Letters, etc.	a	-	-
Prose	-	b	c
Document		[d]	[e]

(Adapted from National Adult Literacy Survey)

Instructions (read aloud by administrator):

Now I am going to show you information from a brochure on household safety. Please read aloud or silently. When you are through reading, look at me or say "Done" to let me know that you are finished. I will then ask you questions about what you have read... Go ahead and read this.

Stimulus (given to respondent)

SAFE WAYS
FOR CLEANING
TOUGH STAINS

Wine: Immediately put salt or hot water on the stain, then soak in milk before washing.

Coffee: Mix egg yolk with lukewarm water and rub on stain.

Chewing gum: Rub with ice: gum will flake off.

1. According to the information, what do you need to clean coffee stains
Correct Response: Must respond both egg yolk and water
2. According to the information, how would you go about safely removing chewing gum?
Correct Response: Must say "rub with ice."
3. According to the information, if you don't have any salt, what else can you put on a wine stain immediately?
Correct Response: Must say "hot water."
4. To clean my wine stained shirt, I immediately put hot milk on the stain, then let it soak in lukewarm water. According to the information, will that take out the stain?
Correct Response: No.

Scoring

Note: Set a mastery level of total to answer correctly to pass this level, then empirically validate with internal consistency methods.

Comments and Guidelines

Some test development guidelines for document tasks:

- select a common print type (announcement, advertisement, brochure) that describes a content and context of high cultural familiarity
- check that documents use familiar, oral vocabulary terms to most native language speakers
- order documents based on complexity of information from simple (e.g., a two column table) to more complex (e.g., a bus schedule with multiple columns and footnotes for interpreting special symbols.)
- make tasks as realistic to the typical purposes and uses of the document as possible (e.g., look up the time a bus leaves; determine dosage on a medicine label.)

SUB-ANNEX 2.2: Writing Assessment Samples

Sample A - Prerequisite level writing tasks

	TYPE OF SKILL			
	BASIC MOTOR	CONVENTIONS	COMPOSITION	APPLIED
DOMAINS				
Letters, etc.	[a]	b	-	-
Prose	-	c	d	-
Document		e	f	g

Instructions (read aloud by administrator):

Please write your name on this line.

Now I would like you to write or print some letters for me. Listen carefully and put the first letter I say here (point). The rest of the letters go in the spaces across this line.

A B R T C M O H P E F L U Y V

[Allow 5 seconds for each letter; if individual is writing when time elapses, allow them to continue.]

I am going to read some words to you and I would like you to write or print them in the numbered spaces. Try to spell them correctly. I will say the word, then read a sentence with the word in it, and then say the word again. Try your best. If you are not sure how to spell a word, it is OK to take a guess.

(phonetic alternatives one might accept)

and
in
him
make mak
cook cuk
enter entir/inter/antr

Now I am going to read some common phrases and I would like you to write or print them in the numbered spaces. Try to use spacing and punctuation like you would read it. I will read the sentence once through, then repeat it again word by word as you write. Try your best. If you are not sure how to spell a word or write, it is ok to take a guess.

Try your best.
Turn left.
Employees must wash hands.

Now I am going to read some sentences and I would like you to write or print them in the numbered spaces. Try to use spacing and punctuation like you would read it. I will read the sentence once through, then repeat it again word by word as you write. Try your best.

My name is < >.
What time is it?

Where is the post office?

I am going to the market to buy eggs, milk, and cheese.

Stimulus (given to respondent)

Your name

Letters

1. ____ 2. ____ 3. ____ 4. ____ 5. ____ 6. ____ 7. ____

Words

1. _____
2. _____
3. _____

Phrases

1. _____
2. _____
3. _____

Sentences

1. _____
2. _____
3. _____

Scoring

Continue until 10 in a row are spelled incorrectly using phonetic alternatives.

[Note: the WRAT has developed a norm-referenced score key for judging levels. Country-specific mastery levels can be determined by a combination of content experts and pilot testing against comparison groups of known ability levels.]

Comments and Guidelines

Some test development guidelines for pre-requisite skills:

- order letters by frequency of use in language (if known)
- select words that appear in print frequently
- select words that are familiar oral vocabulary terms to most native language speakers
- vary parts of speech (verbs, nouns, prepositions) and use most common usage (present tense, non-plural)
- order words in increasing complexity by features such as length, number of syllables, decodability, abstractness

Alternate Item Formats

Print text on answer sheet and have individuals copy onto lines below text. Basic motor fluency can be judged based on accuracy and speed of copying.

Sample B - Basic level writing tasks: prose

	TYPE OF SKILL			
	BASIC MOTOR	CONVENTIONS	COMPOSITION	APPLIED
DOMAINS				
Letters, etc.	a		-	-
Prose	-	[c]	[d]	-
Document		e	f	g

Adapted from Mauritius Exam Syndicate (1996).

Instructions (read aloud by administrator):

Write answers to the following questions in full sentences.

Stimulus (given to respondent)

1. How old are you?

Correct Response: I am ___ (years old.)

2. How many brothers and sisters do you have?

Correct response: I have ___ brothers and ___ sisters.

3. What is your favorite meal?

Correct response: I like eggs and bacon.

4. I am a stranger in this area. Please write directions for getting to your home from ___?

Correct response:

0 - incomprehensible

1 - one or more clear statements using travel conventions (turn right, walk to the fence post, etc.)

2 - detailed, step by step statements

5. Write a list of five or more items to buy at the market?

Correct response: comprehensible list of items appropriate for the market.

6. Write a brief description of what you did today?

Now read all that you have written back to me. Is there anything that you would change if you could?

Scoring

0 - cannot read their own writing or makes many errors

1 - can read their own writing, but shows no awareness of mis-spellings, inappropriate print conventions, left out words, ambiguous or incomplete sentences or steps.

2 - shows awareness of one or more mis-spellings, inappropriate print conventions, left out words, ambiguous or incomplete sentences or steps that may confuse the reader.

Comments and Guidelines

Some test development and administration guidelines for writing skills:

- read all directions aloud in case individual has difficulty reading
- cross-check that respondent understands task and questions by having him or her answer orally after trying to write responses

Sample C - Basic level writing tasks: document

	Type of Skill			
	BASIC MOTOR	CONVENTIONS	COMPOSITION	APPLIED
DOMAINS				
Letters, etc.	a		-	-
Prose	-	c	d	-
Document		[e]	[f]	[g]

Adapted from National Adult Literacy Survey

Instructions (read aloud by administrator):
Fill out the following form.

Stimulus (given to respondent)

Employment Application			
Name _____		Date: _____	
First	Surname	Day	Month Year
Address _____			
City: _____		State/Province _____	
Date of birth: _____			
Day		Month Year	
Availability			
Total Hours Available per week _____			
Employment History			
Two most recent jobs:			
Dates	Company	Supervisor	Reason for leaving
From: _____	To: _____	_____	_____
From: _____	To: _____	_____	_____

Scoring

A scoring guide should be developed that a) establishes guidelines for each fill-in, b) weights items based on the difficulty and importance of accuracy in completing each.

Comments and Guidelines

Some test development and administration guidelines for document writing skills:

- read all directions aloud in case individual has difficulty reading
- cross-check that respondent understands task and questions by having him or her answer orally after trying to write responses

SUB-ANNEX 2.3: Numeracy Assessment Samples

Sample A - Pre-requisite level numeracy tasks

DOMAIN	TYPE OF SKILL				
	DECODING	WRITING	COMPUTING	LOCATING	MEASURING GRAPH
Written numbers	[a]	b	c	d	e
Numbers in oral text	--	--	[f]	--	--
Numbers in prose text	--	--	f	g	h
Numbers in documents	--	i	j	k	--
Everyday objects & actions	[a]	--	--	--	l

Instructions (read aloud by administrator):

1. Show me FIVE fingers
2. How many birds do you see in this picture?



3. Look at the numbers on this line (point). Read me the numbers one by one.



4. You have 6 eggs. Your hens lay 3 more eggs. How many eggs do you have now?

Comments and guidelines

- § These questions examine ability to understand simple number words, count concrete objects, decode simple one-digit or two-digit numbers, and do very simple mental addition.
- § objects and other quantitative elements should be converted into culturally relevant and meaningful units/elements.
- § Respondents who cannot answer most such questions correctly, when presented in their own mother tongue, will appear to not have even the lowest pre-requisite skills and may be classified as "non-numerate". However, countries will have to decide if familiarity with "formal" number systems (such as with Roman numerals above) is part of the threshold skills distinguishing between non-numerate and prerequisite levels of numeracy. It is expected that some individuals will be able to function at a higher level within an indigenous number-system.

Sample B - Prerequisite level numeracy tasks

DOMAIN	TYPE OF SKILL				
	DECODING	WRITING	COMPUTING	LOCATING	MEASURING GRAPH
Written numbers	a	[b]	[c]	d	e
Numbers in oral text	--	--	f	--	--
Numbers in prose text	--	--	f	g	h
Numbers in documents	--	[i]	[j]	k	--
Everyday objects & actions	--	--	--	--	l

Instructions (read aloud by administrator):

1. Look at the numbers on this line (point). Read me the numbers one by one.

10	24	90	110	2001	12.50
----	----	----	-----	------	-------

2. Now write each number I say to you:

6	2	4	12	40	65	125	1995	5.5
---	---	---	----	----	----	-----	------	-----

3. You have 12 cows. You sell 6 cows. How many do you have left?
4. 10 [object] cost 10 dollars. How much do 5 [object] cost?
5. A woman goes to the market to buy flour. She has twenty dollars. Each package [kilogram] of flour costs 3 dollars. How many packages [kilograms] can she buy?

Comments and Guidelines

- These questions illustrate a range of difficulty levels within the pre-requisite level of numeracy.
- oral and written questions can be mostly the same except for presentation. Oral questions should be brief to reduce memory load.
- convert all objects, money, and other quantitative concepts into culturally relevant and meaningful units.
- the complexity of pre-requisite level oral and written math problems must be determined by cultural and economic needs.
- In general, problems become harder as multiple operations are involved, as more complex numbers (e.g., multi-digit, decimals) are involved, as extraneous information is included (i.e., information that is not needed to solve the question), and as the task involves more reading and text comprehension.

Sample C - Basic level numeracy tasks: numeracy/literacy

DOMAIN	TYPE OF SKILL				
	DECODING	WRITING	COMPUTING	LOCATING	MEASURING GRAPH
Written numbers	a	b	c	[d]	e
Numbers in oral text	--	--	f	--	--
Numbers in prose text	--	--	f	g	h
Numbers in documents	--	i	[j]	[k]	--
Everyday objects & actions	--	--	--	--	l

Stimulus (shown to respondent)

Office supplies Ltd.			
ORDER FORM			
Item	Quantity	Unit price	Cost
Paper pack	1	10.00	10.00
Calculator	2	15.00	
Total purchase			

Instructions (read aloud by administrator):

Look at this order form and complete it.

1. Write how much will be the cost of the calculators.
2. Write what will be the value of the total purchase.

Comments and Guidelines

- The item illustrates a work-related functional task. The respondent has to do simple computations (addition, multiplication or repeated addition) and fill a form.
- Money amounts and other quantitative and visual elements have to be converted to be culturally relevant.
- The complexity of such tasks must be determined by cultural and economic needs.

- In general, problems become harder as multiple operations are involved, as more complex numbers (e.g., multi-digit, decimals) are involved, as extraneous information is included (i.e., information that is not needed to solve the question), and as the task involves more reading and text comprehension.

Sample D - Basic level numeracy tasks

DOMAIN	TYPE OF SKILL				
	DECODING	WRITING	COMPUTING	LOCATING	MEASURING GRAPH
Written numbers	a	b	c	[d]	e
Numbers in oral text	--	--	f	--	--
Numbers in prose text	--	--	f	g	h
Numbers in documents	--	i	j	[k]	--
Everyday objects & actions	--	--	--	--	l

Stimulus (shown to respondent)

Children's Aspirin			
Age:	Under 2	2-3	4-6
Weight:	less than 15 lbs	(15-25lbs)	(over 25 lbs)
Dosage:	(see your doctor)	1 tsp	2 tsp
	-	(.8 ml)	(1.6 ml)

1/2 dropper = 1 teaspoon 1 dropper = 2 droppers

Directions: Dosage can be administered every 4 hours up to 4 times a day, as needed. Enclosed dropper holds .8 ml when filled to the line.

Instructions (read aloud by administrator):

Look at this medicine label for Children's Liquid Aspirin.

1. Show me (circle) where you see .8 millilitres
2. Show me (circle) the dosage for a 20 pound 3 year old.
3. How much does the dropper that comes with the Aspirin hold?
4. You have a 4-year old. How much Aspirin should be given?

7.1 Comments and Guidelines

§ This item, similarly to the previous one, requires locating information in a document. It involves the joint activation of reading skills and low-level numeracy skills (decoding of simple numbers, except for some decimal numbers as in "0.8").

§ *The task does not involve computation but only or mostly interpretation of written text and numbers. Correct performance on this task depends more on reading skills than on high-level numeracy skills. The need to combine both skills and locate information within a somewhat complex document contribute to the difficulty level of the task.*

§ *The task illustrates how the same stimulus material can serve as a basis for creating several different tasks, as well as highlights that literacy and numeracy sometimes overlap.*

Sample E - Advanced level numeracy tasks

DOMAIN	TYPE OF SKILL				
	DECODING	WRITING	COMPUTING	LOCATING	MEASURING GRAPH
Written numbers	a	b	c	d	e
Numbers in oral text	--	--	f	--	--
Numbers in prose text	--	--	f	g	h
Numbers in documents	--	i	[j]	[k]	--
Everyday objects & actions	--	--	--	--	l

Instructions (read aloud by administrator):

Here is a distance chart. Calculate the total number of kilometers you will travel in a trip from Guadalajara through Tecoman to Zamora.

Manzanillo	322			
Tecoman	340	273		
Puerto	269	62	330	
Zamora	171	342	515	289
	Guadalajara	Manzanillo	Puerto	Tecoman

Note: This task was created on the basis of an item used in the International Adult Literacy Survey and published in the IALS 1996 report.

7.2 Comments and Guidelines

7.3 This task requires the respondent to perform a single addition operation, but he needs to locate the required numerical quantities in a non-trivial document and in the presence of multiple distracters. Further, a higher level of comprehension is required to understand the meaning of the question (which relates to a trip with two parts). Hence, the task is harder than filling the task involving filling an order form which was presented earlier.

Scoring

Note: Proficiency scales are highly suspect in life skills sub-domains. However, one may interpret a low score percentage of basic items to suggest that an individual is at higher risk of, for example, engaging in behaviors that contribute to poor health and nutritional standards. This technique may not work as well for everyday functional knowledge and procedures like telling time on a clockface or direction from a compass. In either case, interpretive judgements must be validated against other behavioral and practical research information collected on the target group.

Comments and Guidelines

Some test development guidelines for document tasks:

- select knowledge and behaviors that are identified as important life skills issues by community-based standards
- develop distracter sets for multiple choice questions based on common attitudes, beliefs, practices or misconceptions observed in the community (e.g., belief that HIV is transmitted by casual contact)
- weight items that have higher risk potential (e.g., reading a label that says "danger poisonous if swallowed" versus "may cause some stomach irritation")
- make tasks as realistic to the typical life skill problems as possible (e.g., determine dosage on a medicine label.)

Sample B - Basic level life skills tasks

	TYPE OF SKILL			
	KNOWLEDGE	ATTITUDES, BELIEFS, DISPOSITIONS, ACTIONS	CRITICAL AWARENESS	PROBLEM SOLVING
DOMAIN				
Health			[c]	[d]
Everyday life				
Social environment				
Natural environment				
Workplace				

* Each Domain is independent of the others, therefore basic levels cannot be generalized across domains.

Instructions (questions read aloud by administrator; answers can be oral as well, if poor reader.)

1. What should you tell a friend who smokes about the risks of smoking to get him or her to stop?

Scoring:

- 0 - No response
- 1 - Simple slogans like - Don't smoke it's bad for you.
- 2 - One or more health risks (causes cancer) and one or more persuasive comments (it makes your breath smell bad)

2. Many people get sick with dysentery in the summer. Name two things you could do to reduce your risk of getting dysentery?

Scoring:

- 0 - No response
- 1 - Simple slogans like - Wash my hands
- 2 - Two or more healthy behaviors (wash hands after every meal) or two or more action-oriented behaviors (clean the environment)

Comments and Guidelines

Some test development guidelines for life skills:

- oral questions and responses should be used whenever reading skill of the respondent is in doubt.
- survey style items may be most effective for finding out actual beliefs, values, attitudes, and behaviors, but a trustworthy interviewer will be necessary to get answers to sensitive questions.
- the administrator can ask questions as free-response then use distracter list (plus an Other category) to have a list of the most common responses record responses. (see below); this requires uniform, well-trained field staff.
- develop multiple-choice distracters based on actual field-test free responses. For example, if many people have a misconception that

malaria is caused by spoiled food, then a distracter that draws their attention to that choice is helpful.

Alternate Item Formats

[Rather than creating all direct measurement format items, create open-ended survey items asking for direct behaviors and actions as illustrated below.]

1. Do you smoke?
2. What do you usually eat for breakfast?
3. How often do you eat the following: Answer with daily, weekly, less than once a week.
 - Bread
 - Meat
 - Candy/sweets
 - Fried foods

ANNEX 3:

CASE STUDIES OF PREVIOUS HOUSEHOLD SURVEYS WITH COMPONENTS SIMILAR TO THE BLC MODEL

Household surveys, which have utilized measurement of basic learning skills of one or another kind, have been quite varied in quality, intent and policy purposes. For example, some have had multiple policy purposes - ranging from international and national levels to grassroots program accountability studies. Thus, designs for these surveys can take quite varied forms. Drawing on previous work, but extrapolating as well on certain aspects to highlight certain issues, provided below are several case studies that illustrate distinct approaches to sampling, survey development, test development, human resources, and capacity-building.

Case study 1: India: Implementing sampling and quality data collection designs for a national literacy campaign evaluation. Reference: Khan, N. U. (1999). Total literacy campaign of District Bikaner: External evaluation report. New Delhi: Department of Social Work, Jamia Millia Islamia.)

This case study is based on an evaluation report of the National Literacy Mission in Bikaner District of India (Khan, 1999). The literacy campaigns are non-formal educational programs described as area-specific, time-bound, volunteer-based, cost-effective, and outcomes-oriented. They include programs for youth ages 9-15 years, as well as adults ages 16-35 years, thus overlapping the target population of the BLC surveys. The evaluation design is especially informative in the context of BLC surveys for the decision-making process concerning sampling strategy, community involvement and participation, and the quality of data collection processes.

1. Targeting key policy objectives to guide study design

To achieve the main objective of assessing the effectiveness of the literacy campaign programs in the target group, the evaluation team targeted four key policy-based objectives:

- (a) assessing coverage, that is, the extent to which identified illiterates in the district enrolled in the program;
- (b) achievement levels of program participants;
- (c) feedback on strengths and weaknesses of the campaign
- (d) impact of campaign on community involvement and social awareness.

These objectives required a design strategy for sampling learners (both current and completers) who attended programs, as well as verifying the accuracy of the district attendance statistics.

2. Designing and implementing the sampling strategy

National census data was used to calculate overall illiteracy rates by region, villages, and wards for a target population defined as the age group from 9-35 years. District level records were available on total enrollment in the program, total current participants in each of the three instructional levels, completion rates, ages, and genders category wise for all blocks, villages, and wards in the district.

A sample size was calculated to achieve an acceptable level of reliability of estimates of the target population and subdomains, then a multi-stage sampling design was implemented. Villages and wards were considered sampling units. They were listed serially, then randomly sampled under a side condition of proportionate sampling of rural and urban areas. The selected villages were of different population sizes and spread all over the district. Three contingent villages from each block were randomly selected in advance. All villages and wards were divided in two equal sub-samples to test for sampling error.

The evaluation teams also visited villages other than those in the sample, to verify the district level data. A problem evident in previous evaluations was the presence of "proxy" learners, that is, individuals who had not participated in the campaign who showed up at the testing sites as if they had. Through a combination of participatory involvement by local administrators and other quality checks, the evaluators were better able to reduce the number of proxy learners and report separate data about their learning achievement in the Bikaner district study.

3. Building local capacity using a participatory approach

The staff of the local campaign were invited to participate in the process at the time of the drawing of the sample. The sampling process was explained and then the campaign staff helped by making logistical arrangements at the village/ward level. Evaluation field teams were accompanied by at least two representatives of the campaign staff to facilitate the process. The campaign staff advised on scheduling, made boarding and lodging arrangements, provided stationary and carrying bags for evaluation team, and ensured easy access to information. Campaign staff also helped identify "proxy" learners and to check marking of tests (though administration and scoring was carried out by evaluation staff.) In general, campaign staff proved helpful in maintaining a congenial atmosphere and in adhering to the lengthy daily schedules.

In addition, survey interviews were developed and carried out with local district officials, program administrators, teachers, and volunteers to gauge the community awareness and involvement with the literacy program.

4. Recruiting and training human resources to ensure the quality of data collection

The evaluators had considerable experience working with disadvantaged sections of the population and considered it essential to select team leaders and test administrators who had similar experience and understanding. The criteria for selecting evaluation team members included: a) an objective and sympathetic attitude, b) the ability to deal with any

situation with due respect to human attitude and emotion, c) the ability to work as a facilitator and enabler to find out what participants know rather than what they do not know, and d) the capability to politely and firmly handle the problem of 'proxy' learners.

Post-graduate university students were recruited as test administrators. The participation of local campaign staff (see #3 above) further strengthened the credibility and quality of the evaluation team field efforts. Staff training included familiarization with the various tests and scoring systems, the testing situations, and the likely problems to be encountered before and during test administration. A code book was also provided, detailing processes to ensure uniform administration, as well as providing a checklist of do's and don'ts of field work.

In order to be prepared for upcoming issues and problems, nightly staff meetings were organized to submit reports, discuss experiences and issues, and generally reinforce enthusiasm. Daily meetings between field staff and campaign staff also facilitated better understanding and cooperation. Field staff were responsible for randomly rechecking test papers after field work was completed.

Comments: This case study illustrates two innovative features especially relevant to BLC survey designers. First, by building partnerships with the staff of the local literacy campaign, the evaluation team was able to take advantage of their local experiences and understanding in working with the target populations, as well as helping these local literacy providers to develop new capacity for future evaluation efforts. Second, while sampling was done based on random selection of households as identified by census data, testing was actually performed at a central location in each village/ward. Logistically this was possible without threatening the quality of the data collection, because of the strong, positive, long-term community involvement of the literacy campaign staff in organizing and recruiting participants to show up at the test sites.

Case study 2: Developing a numeracy scale for the International Life Skills Survey.

Reference: (ILSS, Iddo Gal, personal communication, 1999)

The International Life Skills Survey (ILSS) is a comparative survey jointly developed by Statistics Canada and by the United States' National Center for Education Statistics, in cooperation with the Organisation for Economic Cooperation and Development (OECD). ILSS will test large nationally representative samples (N=7000 and over) of adults aged 16 to 65 in ten countries in the year 2002. Tasks will assess performance in several domains, including Numeracy, Literacy, Problem Solving, and Practical Cognition; other variables will be assessed via self-report scales and a background questionnaire. This household survey will interview respondents in their homes by trained interviewers.

This case study describes key issues encountered by the ILSS numeracy team, while it worked to develop a scale for assessing functional numeracy skills. One feature of ILSS is that it involves testing in multiple languages. Hence, issues regarding translation and adaptation of items to be fair, valid, and practical for assessment in several cultures should be of interest to countries which plan to survey several language or ethnic groups.

The numeracy team, like other teams within ILSS, has worked through six stages, as described below.

Stage 1: General planning

Three key questions of interest were:

(a) What is the timeline and date of actual implementation in the field? The team developed a timeline, working back (in reverse) from the planned date of implementation of the survey, and estimated time needed for the remaining 5 main stages of work, allocating spare time for delays and unexpected problems.

(b) How many minutes of testing time will be available per person? Numeracy testing was to consist of enough items for 60 minutes of testing (30 minutes per person, but 2 parallel forms).

(c) What do decision makers want to know? Decision makers in participating countries want to know what percentage of people are at different levels of numeracy, including the lowest skill levels (as these usually are the target for further educational efforts). Hence, items span the full difficulty range from very easy to difficult.

The answers to these and related questions helped the team learn about the general study design, context of assessment, and testing framework within which the items and scale developed by the team will be used.

Stage 2: Development of a conceptual framework

Strategic decisions were made in three areas:

(a) Design of a conceptual definition and "road map". What is the nature of the skill domain (i.e., what do we mean by "numeracy" in the context of this survey?) and what facets (e.g., knowledge of four basic operations, percents and fractions, geometry) of numerate behavior are critical to adults?

(b) Deciding what facets of numeracy should be covered with which test items. For example, some aspects of geometry or algebra that seldom come up in everyday numeracy tasks will not be addressed in test items, even though being part of the conceptual definition and part of most schools' curricula.

(c) Decide what is the nature of the tasks and scoring. First, the team had to decide whether to use "school-like" tasks or "functional" tasks. The team decided that all tasks will start from realistic stimulus material (e.g., workplace forms, advertisements, health brochures, food labels, etc), even if it limits the type of mathematical skills that can be tested, as the test has to represent what real people have to cope with in real life. Second, the team considered partial credit important (i.e., give a fraction of a score to show that a respondent knows what to do but makes some computational error). This has implications for training interviewers and scorers. Third, the team had to decide whether people will be allowed to use any tools. Based on the conceptual definition of numeracy, the team decided to provide a calculator, a measuring cup, and a ruler. Interviewers will carry with them a "toolbox" with such items and will make it available to all respondents. (Whether or not they will use these tools is up to the interviewers).

Stage 3: Development of a preliminary item pool

This stage involved many questions and pragmatic decisions. Some of them were:

(a) How many preliminary items to develop for 60 minutes of testing? Strategically, should the test consist of many "short" items, each of which can be answered quickly, or should it consist of fewer bigger problems that require multiple steps? Clearly, the more items the better the coverage of all facets of numeracy. Yet, short items may not represent facets of numeracy that cannot be assessed with short tasks. The decision was to include a mix of both types of items, but use a larger percentage of short items to cover more facets of numeracy. The team estimated that about 15-20 items can be administered in 30 minutes, so at least 80 items would be developed for the preliminary item pool, knowing that some will prove problematic and will be deleted.

(b) What assumptions can be made about the sophistication and training level of the interviewers and the scorers? Highly trained interviewers can use pre-determined probe questions to "extract" more information after an initial response is given by a respondent. Highly-trained personnel can also deal with more complicated scoring schemes

(e.g., that employ multiple codes). However, uniform levels of training in all countries participating in ILSS could not be assumed and therefore relatively simple item formats were developed.

Item development was divided among team members. Each member had to propose not only a set number of items to cover specific facets of numeracy, but also define a coding scheme (i.e., describe what counts as "correct" and "incorrect" answers for each item being proposed), and also had to rate each item in terms of its anticipated level of difficulty (or complexity). An internal review process was then initiated, and based on comments from other team members, items were revised.

Stage 4: Pilot testing of items

This stage involved two key steps:

(a) Informal piloting: team members administered some items that appeared more complex or potentially problematic to a small number (4-8) of respondents. Based on the observed performance and suggestions from the respondents (who knew they were taking part in piloting), items or instructions were revised.

(b) Feasibility study: this study involved administration of items to samples of 300 respondents in each of two countries (U.S. and the Netherlands). For cost reasons, this study was done in a group format, i.e., booklets were administered to groups of people, who answered them in writing. Samples were recruited from diverse sources chosen to represent different age and education groups and were also balanced by gender. The feasibility study had four different but related goals:

(1) Identify problematic items, e.g., items which confuse respondents or are not answered using the mathematical processes which they were designed to capture.

(2) Learn if items turn out to be as easy or difficult as predicted during initial item development.

(3) Test the preliminary scoring system and the extent to which acceptable levels of inter-rater agreement are reached with the scoring instructions.

(4) Examine problems encountered when the items and scoring instructions are translated into another language.

To reach decisions regarding the above goals and decide which items are "good" enough to enter the final item pool, both quantitative (results of various statistical analyses) and qualitative information sources (examining answers to a questionnaire that solicited comments from respondents -- comments by scorers regarding items where there were disagreements with the scoring instructions, suggestions made by test administrators and interviewers, etc.) were used.

Stage 5: Refinement and finalization of the item pool

This stage involves three steps:

(a) Item revision based on results from the feasibility study.

(b) External review. The item pool was submitted for review by selected individuals from countries other than those of the numeracy team members, in order to identify items that may not fit the contexts of other cultures.

(c) Finalization. Based on all commentary and suggestions, a final item pool was created by selecting the best items out of the existing ones and modifying some items.

Stage 6: Field-testing and preparations for full implementation

(a) Testing items and scoring manual: The final item pool will be field-tested in a large-scale "pilot" study before it is implemented in the main study in 2002. This field test will include administration of a total of 90 minutes worth of numeracy items in a household format, to samples of about 1200 respondents in each of the 10 participating countries. Prior to this field study, all items, training materials, and scoring guides will be translated by teams from each participating country and the quality of the translation checked by an independent ILSS translation team.

(b) Testing all logistical elements: The large-scale field-test serves multiple purposes. It enables evaluation of the performance of items as well as all logistical elements of the ILSS survey under the actual conditions expected during implementation of the main study. Key components include: training workshops for interviewers and scorers, quality assurance procedures to ensure consistency of interviews, comparisons (cross-scoring) to check for inter-rater agreement in all languages, reliability of data-inputting (i.e., data keying), etc.

(c) The original 60 items will be reduced to 40 items based on several factors, such as statistical analyses of psychometric properties of the items and their validity or discriminative power, the ability of the items to cover all levels and facets of the conceptual framework developed to define numeracy, and ease and reliability of scoring.

Comments: This case study highlights key elements of the item and test development process used in the ILSS. Although many details will differ, the logic and process are likely to be similar in planning other BLC surveys. In the ILSS process, the feasibility study was a key step that provided much useful information not only on items but also on the quality of data-gathering and scoring processes. Feasibility studies may be a key step in the item development process in BLC-related surveys as well, as they can improve the reliability and validity of the test scores generated and reported by the survey.

Case Study 3: Bangladesh: A low-cost, iterative approach to developing direct measures of reading, writing, written mathematics and oral mathematics.

(Reference: Greaney, V. Khandker, S. R. & Alam, M. (1999). Bangladesh: Assessing basic learning skills. Washington, DC/Dhaka: World Bank.)

This case study is based on a study carried out in Bangladesh to provide baseline achievement information against which future changes in achievement could be measured (Greaney, Khandker, & Alam, 1999). One goal of the study was to remedy the lack of validity in information collected via self-report surveys by using direct learning measures. As no sufficient standard tests in the language were available and imported tests were judged to be of dubious validity, a rapid test development process was undertaken. The Bangladesh study illustrates a low cost, iterative test development process of trial and pre-testing to create direct measures of basic learning competencies, typical of those recommended for BLC surveys. Although operating with less time and resources than available for the ILSS Study (Case Study 2), there are a number of similarities in the key elements of the item and test development decision-making process.

1. Developing definitions and conceptual frameworks

Definitions of basic learning skills were developed, in part, based on "basic learning needs" from the 1990 Jomtien EFA conference (see main text), and through discussions with staff members from the Bangladesh Institute of Developmental Studies and local educators. Basic learning skills (or competencies) were regarded as a continuum of performance levels ranging from non-literate to basic or minimally-competent literate. These definitions were reviewed and approved by an external panel of educators and employers.

The direct assessments in the Bangladesh study adopted a conceptual framework similar to the one recommended for BLC surveys. Levels were defined for each domain (reading, writing, oral mathematics, written mathematics) and items developed to demonstrate mastery at each level. Thus, although basic learning skills were regarded as a continuum of ability, in order to meet practical time (about 15 minutes per interview) and budgetary constraints, as well as the need to provide high quality, easily understood results to a broad audience of stakeholders, a test design that maximized reliable and valid information was created which was based on three and four mastery levels.

2. Trial testing and pre-testing the tests

Test development was divided into two phases: trial testing and pre-testing.

Trial testing. A trial test was developed prior to the full development process of the pre-testing phase for the purpose of determining the feasibility of the conceptual framework, as well as to provide capacity-building opportunities to local staff in objective test construction and administration. A small team of local teachers working

with World Bank experts with objective test development expertise, developed and administered tests to 37 adults in a local village.

The trial test helped uncover administration problems (the novelty of the tests attracted crowds), the need for oral and written mathematics subtests (people poor at written math were often very good at oral math), the fact that the 15 minute target test length was too short to ensure reliable and valid measurement, shortcomings in test administrator training, and the need for more financial resources to achieve project goals.

After analyzing results to screen out items that were too easy, too difficult, or confusing, it was concluded that the results were sufficiently encouraging to justify further test development and funding for a more broad-ranged set of sub-tests.

Pre-testing development phase. A pre-testing development phase followed in which two weeks of intensive training in test development was provided to project-based research officers and field investigators in Bangladesh. Descriptions of achievement levels were agreed upon and a table of test specifications was developed to guide item writing in each of the sub-tests.

A rapid prototyping process was established. Multiple cycles of pilot test trials were conducted in a local area near the offices of the development team. Ready availability of computers and item analysis software meant that data could be analyzed within hours of collection, and items, administration guidelines, and tests could be revised nearly continuously.

The mastery approach adopted required that a passing criteria be selected for each level. At first this criteria was set by test developers, but pre-testing allowed for a more objective, empirically-based criteria later confirmed in the "testing the test" phase. Although under actual test conditions, individuals with higher education would be given automatic credit for lower level items, and testing would terminate when individuals failed to reach the mastery criteria score of a given level, during pre-testing each participant took all easy item and continued to higher level items until the administrator judged that they were not likely to answer any more items correctly.

Following the pre-testing phase, a 15 member review panel of educators, employers, measurement specialists, personnel manager, and researchers evaluated the test. Internal consistency analyses performed on the full study sample are also reported for each level of the subtests, confirming the mastery-level approach assumptions derived from the conceptual framework.

3. Adapting implementation procedures and study goals based on iterative pre-testing process

In addition to describing the results of the full study conducted on a random sample of the Bangladesh adult population, the report also

documents specific problems the development team had in assessing subject area knowledge and in standardizing administration during the pre-testing phase. For example, considerable effort was devoted to developing items on survival skills, improvement of quality of life, and other life skills domains. Data was collected on a 16 item-subtest. Internal consistency analyses proved inadequate for the study and this sub-test was dropped. In another example, various problems arose concerning appropriate standardized administration and scoring guidelines. Failure in some of these efforts to establish quality control procedures for this topic resulted in almost 50% of the writing samples not being returned after the samples had been graded and coded.

Comments: The Bangladesh study illustrates several key points emphasized in the BLC survey recommendations. First, investments in pre-testing can help control costs by identifying problems before they are implemented on a large-scale. Even though the authors of the Bangladesh study needed to revise their goals regarding life skills assessment, it was better to make this determination during pre-testing rather than after investing in a large-scale implementation. Second, by sharing a discussion of the steps taken, difficulties encountered, and decisions made in the development the life skills test, future efforts into developing direct or indirect measures may be enhanced. Finally, the reporting of the loss of data in this study further substantiates the need for adequate training and quality control procedures in data collection.

Case Study 4: Zimbabwe: Combining a background questionnaire with direct literacy measures in a household survey.

(Reference: United Nations Statistical Office (UNSO). (Principal authors: Wagner, D.A., & Srivastava, A.B.L.) (1989). Measuring literacy through household surveys. (Doc. No. DP/UN/INT-88-X01/10E). New York: United Nations Statistical Office.)

This case study is based on a household survey carried out to collect literacy and numeracy information in Zimbabwe (United Nations, 1989). One goal of the study was to evaluate the reliability and validity of self-report measures against direct measures, as well as to collect demographic, educational history, literacy use and practices information (see also Zimbabwe Survey sample forms in Annex 4). For BLC survey designers, the study design may be especially informative in issues of questionnaire development.

1. Designing the survey questionnaire

Several features of the Zimbabwe survey questionnaire design serve as models for BLC studies. First, the opening survey section (Survey Form, page 1) requires the field agent to provide specific administrative control data to help with later with data entry, screening, and checking for sampling error. The household survey was designed to collect information about all members of the household. Thus, a table format is used to code specific information about each member (and visitors) to the household. In addition to demographics, the specific information includes relationship of member to head of household, citizenship, ethnic group, maternal language, school attendance, reason for not attending school. To assess literacy use, all literate members of the household are asked two questions: the number of books and magazines they possess and the languages of these print pieces.

The survey includes 14 additional questions of Educational Background information (Survey Form, page 2). Specifics about attendance in formal school and adult literacy classes, time spent in classes, the languages used for instruction, and reasons for attending or leaving are all probed.

The survey includes another 36 questions of Language, Literacy, and Numeracy information (pages 3-4). These questions range from direct self-assessments of ability (e.g., Question 16. How well do you speak and understand Shona?), to literacy practices, both personal (e.g., Question 19. Do you ever read in your spare time?) and social (Question 26. Do you ever read anything for people who can't read?), to preferences and dispositions (e.g., Question 24. In what language do you most often write?; Question 25. Which language do you think is most useful for you to read and write?; Question 28. What do you think is the MAIN reason for people to learn how to read?). The response categories and scales for these questions can be correlated to direct measures to better assess the reliability of self-report measures, as well as identify indicators that may be useful in education policy or program planning.

2. Direct assessment tests including what to test and how to conduct the tests

Following the questionnaire is the Direct Assessment score sheet with administration directions for the interviewer. The interviewer is directed to use one form per person and to test each person in three languages. The instructions also specify when the interviewer should provide help and guidance (during Practice items) and when he or she should not help. The direct measures consist of a reading, writing, and numeracy test. Pictures and multiple-choice item formats are used extensively making the scoring easier and more objective for the interviewer.

Comments: This study illustrates two innovative features especially relevant to BLC surveys. First, by including extensive, detailed questions about educational background, literacy and language practices, and attitudes and dispositions towards education, policy makers will be able to better model and plan the kinds of policies that are most likely to meet the perceived needs of the target population. Second, the decision to ask about and assess in multiple languages provides a better estimate of the actual literacy and numeracy skills, as well as helping to inform language planning policies.

ANNEX 4: SAMPLE SURVEY FORM FROM ZIMBABWE HOUSEHOLD LITERACY SURVEY

Note: This document was prepared by the CENTRAL STATISTICAL OFFICE of Zimbabwe in 1986. For complete forms, see annexes in: United Nations Statistical Office (UNSO). (principal authors Wagner, D.A., & Srivastava, A.B.L.) (1989). Measuring literacy through household surveys. (Doc. No. DP/UN/INT-88-X01/10E). New York: United Nations Statistical Office.

RECORD TYPE 1 LITERACY SURVEY 1986

Date of

____/____86

Starting Time: _____

Survey No.	Round No.	Admin . Area	Div.	E.A. No.	Segment No.	Sub-Sample	House -hold	Year	Month	Ecol -Zone	Rec. Type
------------	-----------	--------------	------	----------	-------------	------------	-------------	------	-------	------------	-----------

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----

FOR ALL USUAL MEMBERS OF HOUSEHOLD AND VISITORS

SERIAL NO.	USUAL MEMBERS OF HOUSEHOLD AND VISITORS (WHO STAYED HERE LAST NIGHT) NAME	USUAL MEMBERS PRESENT LAST NIGHT - 1 ABSENT LAST NIGHT - 2 VISITORS- 3	RELATION-SHIP TO HEAD OF HOUSEHOLD (CODES GIVEN BELOW)	SEX M - 1 F - 2	AGE AT LAST BIRTH-DAY	AGE AT LAST BIRTH-DAY
24-25		26	27	28	29	30
01						
03...						
15						

MARITAL STATUS	CITIZENSHIP	ETHNIC GROUP	MATERNAL LANGUAGE	FOR CHILDREN AGES BETWEEN 5-14	CHILDREN AGED 5-14 NOT ATTENDING SCHOOL
Never Married - 1 Married - 2 Divorced - 3 Separated - 4 Widowed - 5	Zimbabwean - 1 Mozambican - 2 Malawian - 3 Zambian - 4 Other African - 5 Other - 6	Black - 1 White - 2 Coloured - 3 Asian - 4 Other - 5	Shona - 1 Wilebele - 2 Venda - 3 Langa - 4 Engliss - 5 Malanga - 6 Shangani - 7 Other - 8 (Specify)	School Attendance <hr/> At school - 1 Not at school - 2	Main reason for the child not attending school (codes given below)
31	32	33	34	35	36

Ask literate members of Household

40. Are there any books and/or magazines in your household? If so, how many?

None - 0	1-3 - 1	4-6 - 2	7-12 - 3	more than 12 - 4
----------	---------	---------	----------	------------------

Shona only = 1	Ndebele only = 2	English = 3	Shona and English = 4
Ndebele and English = 5	Shona and Ndebele = 6	All three languages = 7	Other... = 8 (specify)

RECORD TYPE 2

15 YEARS AND ABOVE

PROVINCE: _____ ADMINISTRATIVE AREA: _____ DIVISION: _____

(EDUCATIONAL BACKGROUND)

1. Have you ever attended formal primary/secondary or correspondence school?

Yes - 1 (Go to Q.2)

No - 2 (go to Q.4)

2. What is the highest level of education you completed?

.....

Enumerator: Write respondent's grade/standard in space above. See Manual, page 12 for correct Education code to enter. (Go to Q.3)

3. How long ago did you leave school?

Still in school - 1

0-12 months ago - 2

1-2 years ago - 3

3-4 years ago - 4

5-9 years ago - 5

10-14 years ago - 6

15-19 years ago - 7

20 or more years ago - 8

4. Are you now, or have you ever attended reading and writing (adult literacy) classes outside of Primary or Secondary school?

Yes - 1 (Go to Q.5)

No - 2 (Go to Q.13)

5. What type of Adult literacy classes have you recently attended?

National Literacy Campaign - 1 (post independence)

ALDR/ALUZ classes - 2

Other- 3 (specify)

6. For how long have you been/did you attend these Adult Literacy classes?

0-6 months - 1

7-12 months - 2

1-2 years - 3

3-4 years - 4

5-7 years - 5

more than 7 years - 6

7. How many days a week did/do you attend these adult literacy classes?

One day a week - 1

Two days a week - 2

Three days a week - 3

Four days a week - 4

Five days a week - 5

Other - 6 (specify)

8. Are you still attending these adult literacy classes?

Yes - 1 (Go to Q.10)

No - 2 (Go to q. 9)

9. If not, when did you last attend such classes?

- 0-6 months ago - 1
- 7-12 months ago - 2
- 1-2 years ago - 3
- 3-4 years ago - 4
- 5-7 years ago - 5
- more than 7 years ago - 6

10. What language, or languages are/were you being taught to read and write in these adult literacy classes?

- Shona only - 1
- Ndebele only - 2
- English only - 3
- Shona and English only - 4
- Ndebele and English only - 5
- Other - 6

11. Why did you want to learn to read and write?

- Writing letter - 1
- Acquiring knowledge-2
- Feels good to read/write-3
- Better communication - 4
- For self-help projects - 5
- Inquiries life/health - 6
- other - 6 (specify)

12. Has a friend or member of your family tried to help you to improve your reading ability?

- Yes - 1 (go to Part R)
- No - 2 (Go to Part R)

13. What is the main reason why you are NOT attending adult literacy classes?

- Already literacy - 1
- Lack interest - 2
- Lack time - 3
- Teacher/classes not available - 4
- Drinking beer-5
- Won't mix sexes-6
- dropped out - 7 (Go to Q. 14)
- Other-8 (specify)

Note: (for all responses) except "dropped out" , go to part B of the questionnaire. If respondent says he/she dropped out, go to Q. 16).

14. What is the main reason why you dropped out?

(if respondent who dropped out of adult literacy classes)

- lack of interest - 1
- lack of time - 2
- don't like classes - 3
- teacher/classes not available - 4
- too shy/embarrassed - 5
- drinking beer - 6
- won't mix sexes - 7
- too old - 8
- other-9 (specify)
- Go to Part B of questionnaire)

Record type 3

Province _____ Administrative Area: _____ Division: _____

B. LANGUAGE, LITERACY AND NUMERACY BACKGROUND

15. What is your Maternal (best) language?

- Shona - 1
- Ndebele - 2
- English - 3
- Other - 4 (specify)

(Note: Ask Q.16-28 in Maternal language.
First, ask Q.16-18 about Maternal language: then ask Q.16-18 about English; then ask Q.16-18 about second vernacular. Complete other only if respondent says he speaks another language).

16. How well do you speak and understand?

- very well - 1
- fairly well - 2
- very little/not at all - 3

17. How well can you write in?

- Very well - 1
- Fairly well - 2
- Very little/not at all - 3

18. How well can you write in?

- very well - 1
- fairly well - 2
- very little/not at all - 3

(Enumerator: only ask respondents who say they can read a newspaper in any language.

19. Do you ever read in your spare time?

- Yes - 1 (Go to Q.20)
- No - 2 (Go to Q.22)

20. (If yes), what do you most often read?

- newspaper/magazine - 1
- novel/story book - 2
- advertising/posters/billboards - 3
- bible/religious materials - 4
- other - 5 (specify)

21. What language do you prefer to read this material in? (Codes 1-8)

22. Have you ever written a letter?

- Yes - (Go to Q.23)
 - No - (Go to Q. 25)
-

23. (If yes) To whom have you written most often?

- relative - 1
- friend - 2
- boss - 3
- teacher - 4
- government office - 5
- other - 6 (specify)

24. In what language do you most often write? (Codes 1-8)
25. Which languages do you think is most useful for you to read and write in? (Codes 1-8) (Enumerator: Only ask respondent Q.26 if he/she responded very well - 1, or fairly well - 2 to Q. 17 other wise skip to Q. 28):
26. Do you ever read anything for people who can't read? If yes, what do you read most frequently for such people?
- never read for others - 1
 - letters - 2
 - newspaper/magazine - 3
 - novel/storybook - 4
 - bible - 5
 - medicine labels - 6
 - forms/bills - 7
 - racing card - 8
 - other - 9
27. Do you ever write anything for people who can't write? If yes, what do you write most frequently for such people?
- never write for others - 1
 - letters - 2
 - forms/bills - 3
 - work purposes - 4
 - diary - 5
 - bible/religious materials - 6
 - other - 7 (specify)
28. What do you think is the MAIN reason for people learning how to read? (What will they do with it?)
- Read letters - 1
 - Read for knowledge - 2
 - Read bills/forms - 3
 - Read street signs - 4
 - Job requirement - 5
 - Communication/
Information - 6
 - Other..... - 7 (specify)
29. For whom is it most important to read and write: men, or women, or is it the same for both?
- men - 1
 - women - 2
 - same for both - 3
- (Go to Q. 30)
30. Are you a paid employee?
- Yes - 1
 - No - 2
- (Enumerator): If respondent says Yes, go to Q.31. If respondent says no, go to Q.32.
31. In order to get your current job, was it important to know how to read and write?
- Very important - 1
 - Fairly important - 2
 - Not at all - 3
 - Don't know - 4

32. How often do you use reading and writing in your job/occupation (including your household activities)?

- Regularly/often - 1
- Sometimes - 2
- Never - 3

33. (Enumerator): if respondent answered very little/not at all - to Q. 17 then ask Q.33. otherwise, go to 34).

What would you most like to be able to read?

- Government information - 1
- Letters - 2
- Newspapers/magazines - 3
- Books/novels - 4
- Church related materials - 5
- Instructions/labels/signs - 6
- Work related material - 7
- Advertising/posters - 8
- Other - 9 (specify)

34. When you have to do arithmetic (for example, calculating the total cost of items you buy at the market, can you do this yourself, or do you ask someone to do this for you?

- Yourself - 1 (Go to Q. 35)
- Someone else - 2 (Go to C. 36)

35. If you have a difficult arithmetic problem (for example, you buy 13 items at 39 cents each), do you usually add the total cost in your head, or do you usually use pencil and paper?

- In head - 2
- On paper - 2
- Sometimes head/sometimes paper - 3
- Other..... - 4 (specify)

36. Only for head of household (If respondent is engaged in Agriculture)

When you harvest your crop, do you calculate how many bags you keep, how many you sell, and the amount of money you will receive or does someone else calculate this for you.

- Yourself - 1
- Someone else - 2
- No calculations - 3

Go to Test 1 : Word - Picture Matching test.

SCORE SHEET TEST 1: WORD-PICTURE MATCHING TEST

PROVINCE: _____ ADMINISTRATIVE AREA: _____ DIVISION _____

Fill in number of form(s) used for Test 1:

- Form 1 - 1
- Form 2 - 2
- Form 3 - 3

SHONA		ENGLISH		NDEBELE	
test order _____		test order _____		test order _____	
correct	incorrect	correct	incorrect	correct	incorrect
P1	_____	P1	_____	P1	_____
P2	_____	P2	_____	P2	_____
P3	_____	P3	_____	P3	_____

Note: Proceed to next section only if respondent gets at least 2 out of 3 practice items correct.

1	_____	1	_____	1	_____
2	_____	2	_____	2	_____
3	_____	3	_____	3	_____
4	_____	4	_____	4	_____
5	_____	5	_____	5	_____
6	_____	6	_____	6	_____
7	_____	7	_____	7	_____
8	_____	8	_____	8	_____
9	_____	9	_____	9	_____
10	_____	10	_____	10	_____
11	_____	11	_____	11	_____
12	_____	12	_____	12	_____
Total	_____	Total	_____	Total	_____
correct	Total	Correct	Total	Correct	Total
	incorrect		incorrect		incorrect
	incorrect				

(Note: Do NOT count practice items (P1-P3) in these totals.

If respondent gets 4 or more points in a language, or languages, then test respondent in appropriate language or languages on subsequent tests.

NAME: _____ HH NO. _____ SERIAL NO. _____

Instructions:

Use one form for each person. Test each person on all 3 languages. On practice items, it is helpful to ask the respondent to orally name the object first (e.g. ask, "What is this?" then ask, "Where is the word for this object"). Do not request naming of objects on regular items.

You should provide help and guidance only on the Practice items (P1-P1); do NOT provide help on the remaining items. Proceed to Test items (1-1223) only if respondent succeeds in obtaining 2 out of 3 correct on the practice items.

You can provide the correct answers on the Practice items. Do NOT provide the correct answers on Test items; just say "GOOD" and go to next item. Be sure to complete all 12 test items once you start them.

Provide no more than 30 seconds for each Regular item.

SCORE SHEET F (Q TEST 2: ORAL READING TEST

FILL IN NUMBER OF FORM (S) USED FOR TEST 2

	Shona	English	Ndebele
Form 1 -1	_____	_____	_____
Form 2-2			
Form 3-3			
SHONA	ENGLISH	NDEBELE	
Total words	Total words	total words	
Correct	correct	correct	
_____	_____	_____	

Instructions to Enumerators :

Given one (1) mark for each word correctly read.

Only test the respondent in the language or languages in which he/she scored 4 or more points in word-Picture-Matching, test 1.

Enter the total of correctly read words in the appropriate boxes.

Record Type 4

SCORESHEET FOR TEST 3: WRITING TEST

PROVINCE: _____ ADMINISTRATIVE AREAS: _____ DIVISION: _____

Fill in number of form(s) used for test 3.

	SHONA	ENGLISH	NDEBELE
FORM 1 - 1	_____	_____	_____

SHONA Total	ENGLISH Total	NDEBELE Total
_____	_____	_____

Accuracy Scale _____

Instructions to Enumerator :

- Give one (1) mark for each correctly spelt word
- Enter the total of correctly spelled words in the appropriate boxes.
- Remember to indicate which language(s) and form(s) you are using.
- Do not judge the speed or beauty of the handwriting.

¹ This paper benefited greatly from the input of the participants who attended the Expert Workshop (see Annex A). Additional inputs were obtained from Workshop resource persons: Meng Hongwei, R. Govinda, Iddo Gal, Vincent Greaney, Scott Murray, and R. A. Manrakhan. Most of the writing and the numerous revisions contained in this document were prepared by Dan Wagner, Iddo Gal and John Sabatini.

² See Greaney, Khandker, & Alam (1999) for a detailed household survey of youth skills in Bangladesh; also, Wagner (1999) for an overview of literacy issues in developing countries.

³ In Greaney et al. (1999) a World Bank sponsored survey showed that children in Bangladesh who had received up to 3 years of primary schooling were virtually indistinguishable from non-schooled same-aged children in basic skills; this result has called into question the

overall effectiveness of the primary education system.

⁴ See Wagner (1992).

⁵ Surprisingly often, school pupils with five (or even more) years of schooling are found to be functionally illiterate. Further, indirect measures are based on the assumption that people can be classified as either literate or illiterate, and in the same fixed way across countries.

⁶ Others, such as in the IALS project (described earlier) have preferred to use continuous scales to describe performance. As noted elsewhere, policy making rarely requires such specificity as individual scale scores; furthermore, given the past history of bimodal (literacy vs. illiteracy) scaling in literacy work, a four point scale is likely to be easier for practitioners and policy makers to understand.

⁷ In most cases, the adequacy of these four levels will need to be ascertained by national panels of experts, so as to assure that each level contains sufficient breadth to capture the competencies of that portion of respondents. As pointed out later, pretesting will be an important feature of appropriately developing items for these four skill levels.

⁸ Values and attitudes also exist concerning literacy. Every item of print represents cultural and social voices, values, and biases; there is no such thing as a strictly "objective" text. School-based or official documents (e.g., legal contracts) may result in affective or critical responses in out-of-school youth not anticipated. Such critical responses can be viewed as constructive (reading the world) or maladaptive, depending on one's point of view. More work needs to be done in the area of understanding how to measure attitudes and values toward literacy, but this is beyond the scope of the present study (see however, the present discussion of values and attitudes in life skills).

⁹ Though broad definitions of literacy include both reading and writing abilities, these skills are not so interdependent as to prevent an individual from learning to read, but have little or no writing skills. The reverse is less common. However, an individual could have relatively more strength in composing simple, understandable messages, for example, in a second language, than in reading in that language, because reading would require greater command of conceptual vocabulary and linguistic structure. In general, literacy definitions are discussed (below) in terms of the demands of reading first, then writing.

¹⁰ See Gal (1997). Numerate behavior is enhanced when people maintain a comfortable feeling with numbers and when they possess positive dispositions, attitudes, and beliefs, such as about themselves as problem-solvers or about the usefulness of mathematical skills.

¹¹ See Furniss & Baldo, (1998), p. 1.

¹² Greaney, et al. (1999) found that one life skill deemed very "functional" involved survival behavior under flood conditions, a situation which is very common in large parts of this country. However, children who lived at higher altitudes had little relevant knowledge regarding floods at all. In another example, from Mauritius (Mauritius Examinations Syndicate, 1996), a survey of children's life skills included knowledge of a policemen's hand signal for "stop", since policemen are usually positioned outside schools and need to

tell the children when to stop if a car arrives. However, many rural children had little or no experience with crossing streets and avoiding cars; and in some countries, there are no policemen near schools or where they do not use hand signals. Thus, survival skills may vary both across and within cultures.

¹³ Some disadvantages of national, census-based survey methodologies include high costs to implement, limited time to assess individuals, and cost compromises in tailoring to subgroups.

¹⁴ Pilot studies in which individuals or focus groups are surveyed then interviewed about the meaning and fit to questions are critical to avoid missing vital elements.

¹⁵ See ILI-Unesco Paris expert seminar report (1998).

¹⁶ Where schooling is compulsory, access to out-of-school youth may be a politically and socially sensitive issue for local officials, families, and the youth themselves. Youth may not be located as readily in traditional households, as they may be living or working in a city or elsewhere. Thus, identification and estimation of the target population and local context will be a critical need in data collection. Involving local stakeholder groups in the planning of the study (see section 6) will be key in such contexts.

¹⁷ Similarly, asking for specific ethnic or religious group affiliation is preferable to broad categories such as race. This principle must be balanced against the need for efficiency, cost-effectiveness, and reliability of administration, data reduction, and analysis needs.

¹⁸ Each level (up through a continuous scale) requires more data to maintain reliability; thus fewer levels are cheaper in time, money, and expertise. However, the use of three levels (as opposed to four) runs into the problem that evaluators have a tendency to score individuals 'in the middle' when they do not have time or expertise to make finer discriminations of behavior. A four-level scale has been found to be a good compromise for the purposes of this type of approach.

¹⁹ The NALS (Kirsch, et al., 1993) and IALS (OECD/Statistics Canada, 1995) studies illustrate continuous scale tests which employ a threshold level methodology. NALS and IALS defined five proficiency levels based on a continuum of ability, with an 80% proficiency level of items at particular degree of difficulty being the operationally defined threshold. The technical decisions of how and where to define thresholds is critical to policy decision-making. If NALS and IALS had used a 50% item proficiency level (as is sometimes used in other educational measures), then the distributions of abilities among the five levels would have been quite different. If they had chosen only four levels the distributions would also be much different. The NALS and IALS employed IRT (Item Response Theory) which can achieve greater statistical efficiency with fewer test items. IRT is a mathematical technique for predicting the response of an individual of any ability to a single item (test question) of any difficulty. The result of this technique is an "item response function" that shows the probability that a given individual will correctly answer that item. IRT can be used to select the items tailored to the respondent's estimated ability. However, the relative cost of IRT may be high due to the need for highly trained test specialists and statisticians (see Tuijnman, Kirsch & Wagner, 1997).

²⁰ Naturally, difficult items should be answered correctly more often by the most able; easier items should be answered correctly by all respondents. In item selection, it is important to check the correlation of each item with total proficiency scores. Also, local test content developers must have a rationale about what makes items more or less difficult based on the skill domain. For example, in prose literacy items, theory predicts that the complexity of the text, the number of task operations to perform, and the plausibility of distracting information will predict difficulty (Kirsch & Mosenthal, 1990). These features can be manipulated in the task then tested with respondents.

²¹ One could include pre-reading skills such as familiarity with the conventions of reading print, such as recognizing the top from bottom of print piece, the direction to read the text, awareness that the goal of reading is to comprehend or understand the text, etc. Such pre-reading skills are sometimes used for assessing school readiness in beginning readers, and may have value when there are numbers of individuals with some exposure to print (resulting, say, from short-lived literacy campaigns). See Wagner (1993) for such measures used in Morocco.

²² Although comprehension is more often the key skill in "prose literacy" and an application in "document literacy" as operationalized in the Kirsch and Mosenthal (1990) theory, in practice these skills and domains are best viewed as a matrix (see Table 1b).

²³ In English or Arabic, learning to decode (translate print into sound) is a major reading skill achievement that continues to develop even after comprehension and applied skills are in evidence. The distinction is referred to as transparency (simple print to sound translations) versus opacity (complex print to sound relations) in the reading literature; (see Oney, Peter & Katz, 1997) In more opaque languages, reading complex words may be a reliable index of reading ability, because it indicates wider reading experience. By contrast, in ideographic writing systems (e.g. Chinese), the ability to recognize an ever increasing core vocabulary of characters is both a prerequisite skill to advanced levels of reading, as well as a lifelong learning task. Consequently, learning to read Chinese without direct formal or informal instruction may be unrealistic for most individuals. Nonetheless, expecting mastery of thousands Chinese characters as a prerequisite to functional understanding and comprehension of Chinese texts may be equally unrealistic.

²⁴ The distinction was originally proposed by Chall (1983) as distinct stages of reading development. However, as discussed here, comprehension and applied skills are both necessary to basic learning competency in reading. Neither is more developmentally advanced than the other, though they have different contexts of use.

²⁵ Vocabulary or word knowledge is also an important element of reading abilities, but it may be a problematic technique for BLC assessment purposes. Generating word definitions is typically an academic versus natural language ability. Similarly, synonyms and antonyms require linguistic flexibility which is not the same as basic reading competency. Finally, oral or receptive vocabulary knowledge is confounded with indicators of language or verbal ability.

²⁶ The tasks and print examples that define level 2 of document and

prose literacy in the IALS (OECD/Statistics Canada, 1996) may be good candidates for operationalizing advanced levels. First adopted in several large-scale national studies in the U.S., Canada and other countries (Kirsch & Mosenthal, 1990), these surveys argue for the need to distinguish between facets of literacy that involve somewhat different skills and cognitive processes, primarily between *Prose Literacy* (i.e., the knowledge and skills needed to understand information from texts such as news stories, editorials, poems, informational brochures or leaflets, and fiction literature) and *Document Literacy* (the knowledge and skills needed to locate and use information in documents employing different formats, such as forms, tables, maps, charts, job applications, or schedules).

²⁷ Judgments must also be made to the relative importance of print conventions to communication. Are accurately formed letters or punctuation marks essential to understanding a message or can a reasonably skilled reader make sense of the writing without such conventions?

²⁸ Street and store signs, or addresses are good examples of local print. Writing the names of simple object words based on picture drawings or verbal naming are a slightly more complex option.

²⁹ Analogous to the reading assessment scheme, composition and applied skills are both considered necessary to basic learning competency in writing. Neither is more developmentally advanced than the other, though they have different contexts of use.

³⁰ Examples for such informal skill development that are important for economic sustainability have been amply documented in the research literature, such as the mathematical skills of unschooled street vendors in Brazil (Carraher, Carraher, & Schliemann, 1985).

³¹ In any society, individuals may perform numeracy functions on a wide array of materials. Certain individuals may also specialize in specific types of numerate domains (e.g., store-keepers, tailors), and have considerable expertise in mental arithmetic. Even individuals with low general levels of numeracy skill may be able to successfully cope with written materials in a domain in which they have a great deal of practice (e.g., mothers who mix baby formulae or farmers who deal with a certain kind of pesticide). Therefore, since governments and agencies are generally interested in providing numeracy for many categories of people, it is useful to sample tasks across the domains where numeracy functions typically are found. These materials may also appear as single numbers or on signs, labels, texts, advertisements or documents.

Real-life tasks may require both numeracy and literacy skills, and it will be difficult to determine which of the skills (or lack thereof) is responsible for an observed performance level. To illustrate, the sample skills and domains chosen in Table 3c include not only the traditional cluster of computational skills, but also measurement. As also can be seen in Table 3c, some cells may involve a single simple step (e.g., decoding a number on a weight scale), or multiple steps (as in a complex problem solving task). The actual difficulty level of the sample tasks can be varied significantly by changing parameters such as: the complexity of numbers involved (e.g., "7" or "1,609" or "1.2 billions"); the number of operations or reasoning steps called for, and the nature of these operations or steps; the degree to which the problem is amenable to mental calculations (as opposed to necessitating the use

of written or computational aids); and the degree of the respondent's familiarity of the context of the problem.

³² As with literacy, the sample numeracy tasks listed in Table 3c are ordered roughly in terms of increasing level of numeracy skills required. However, not all facets of numeracy fit a simple hierarchical skills structure. The actual difficulty level of the sample tasks described in Table 3b, and more generally of any task, can be varied significantly, by changing parameters of the tasks used. The basic level constitutes some ability to problem solve, that is use skills of computing, applying, and measuring in genuine functional contexts of use. Some countries may require command of various types of prerequisite knowledge and fluency with the formal or print-based numerical, symbolic, and measurement systems. Others may reserve this for advanced levels. The assessment also has to vary the *literacy demands* of numeracy tasks, both to make sure that respondents with low formal literacy or numeracy skills are not penalized, as well as in order to represent the rich array of combinations of literacy and numeracy demands within real-world tasks.

³³ Attitudes, values, and dispositions are descriptors that could be given separate skill categories. They are clustered here in part because they all stand in the same relation to action, critical awareness, and problem solving, that is they interact, but not necessarily in a predictable causal direction. A person can believe, value, or be disposed in one way, then act in direct contradiction because of conflicting values or compromising circumstances.

³⁴ Measurement of life skills can also be confounded with literacy and numeracy abilities, or lack thereof. Thus, written answers will almost always underestimate the life skill knowledge of those with little or no schooling. Individual interviews, in local languages, are thus an essential part of this type of interview assessment.

³⁵ See Chinapah (1997) or Mauritius Examinations Syndicate (1996) for examples of knowledge and skills items.

³⁶ In the development of life skills test items, correlational analyses should be undertaken between survey items and target outcomes, as these will help to validate the assessments (e.g., health and nutrition items, such as the relation between high performance on nutrition items and healthy eating habits).

³⁷ A just released draft set of guidelines for health-related "Life Skills Education for Schools" gives some hope that measurement tools could be based on such a curriculum. However, this curriculum is designed for school children, and not yet available for practical use. Nonetheless, items might be taken and field tested for the development of life skills test items (see WHO, 1999).

³⁸ One concern with sharing testing instruments at the local level is that, in subsequent years, surveys may be done where respondents will have practiced on the survey items. This problem can be resolved by varying the tests (and test items) across temporal and geographical cohorts - always a good practice in any case.

³⁹ National governments which engage an international team of consultants, with little connection to local specialists, are likely to leave little trace of their efforts in local capacity building.

⁴⁰ Another good strategy, with longterm implications, is to recruit

university graduate students in disciplines such as health, education, psychology, language, or anthropology, who might then be influenced to choose a career in educational evaluation.

⁴¹ Because most household surveys will be usually administered one-on-one by a survey administrator, constructed response questions can often be used effectively instead of forced- (multiple-) choice test items (taken down and scored by the administrator). In many cases, asking the respondent for oral responses (in their first language) will reduce ambiguity, decrease time of administration, and increase reliability and validity. This is because developing good, culturally-sound distracters (choices) for forced-choice items is difficult, and because forced-choice, as a technique, may be difficult for non-schooled respondents to master. Forced-choice items can also require more reading and interpreting than is needed for valid measurement, resulting in increased error variance. Also, with multiple-choice, one has to take guessing into account, which adds to the number of items that must be administered. Nonetheless, forced-choice items may be preferred when (a) survey administrators cannot be trained to reliably record and score responses or (b) the efficiency of group administration is required; and where such test items are known, through pretesting, to be reliable.

⁴² Pretesting refers to efforts to try out test or survey items on a few individuals before employing in the field. Pilot studies typically refer to small scale preliminary studies, such as a sample in one locale of the full battery of assessment instruments, before doing such a study on a national scale. In both cases, efforts would be made to analyze results, and recalibrate and revise instruments for the next stage of work. Both pretesting and pilot studies are recommended especially in multilingual settings, sometimes more than once each as necessary.

⁴³ One useful method is to try out implementation strategies under so-called "favorable conditions." Elements of this approach include the use of well-trained research assistants to collect data, assessing easier to reach populations, providing incentives, and so forth. The goal of this approach is to evaluate assessment tools, data processing plans, quality of the data and analysis techniques, and implementation procedures under ideal conditions - mainly to determine if the conceptual and analytic approach is appropriate. If flaws are discovered in the process, they can be corrected before they become confounded with other conditions that arise from "less favorable" conditions.

⁴⁴ In this context, analytic redundancy means that additional items simply repeat the same information previously collected, while providing little in the way of new variance explained.

⁴⁵ Generally speaking, incentives can facilitate participation rates and ease the job of the interviewers by creating interest in the survey. The level of incentive (if one is needed) will have to be judged in terms of local norms.

⁴⁶ Sufficient attention to detail has been shown in many studies to be crucial, as is double-checking of procedures and timely monitoring and cross-checking of accuracy of data entry or scoring of responses. For example, it may be useful to plan to cross-check as much as 50% of the first surveys collected by survey takers (interviewers) before they are authorized to collect more data. To ensure quality control, information

about the examiner and date of survey administration should always be noted on each data sheet, as well as a decision matrix based on different types of abnormalities found during the screening process. A useful strategy is to have data collectors keep daily log notes of problems encountered, and then meet daily with supervisors to discuss and resolve them.

⁴⁷ This report is not, however, designed to provide a complete and comprehensive step-by-step methodology of all that is required for such surveys. Rather, it was felt that when such surveys have been implemented, the reports emanating from them will provide the best hands-on detailed examples, and will at the same time be embedded in the context of concrete assessment activities.