

Background paper prepared for the
Education for All Global Monitoring Report 2006
Literacy for Life

Monitoring and measuring literacy

Daniel A. Wagner
2005

This paper was commissioned by the *Education for All Global Monitoring Report* as background information to assist in drafting the 2006 report. It has not been edited by the team. The views and opinions expressed in this paper are those of the author(s) and should not be attributed to the *EFA Global Monitoring Report* or to UNESCO. The papers can be cited with the following reference: "Paper commissioned for the *EFA Global Monitoring Report 2006, Literacy for Life*". For further information, please contact efareport@unesco.org

July 5, 2004

Daniel A. Wagner
International Literacy Institute
University of Pennsylvania

Author's address: Daniel A. Wagner, Professor and Director, National Center on Adult Literacy/International Literacy Institute, University of Pennsylvania, 3910 Chestnut St., Philadelphia, PA 19104-3111. Email: wagner@literacy.upenn.edu

Acknowledgements

This paper was commissioned by UNESCO as a contribution to the 2005 EFA Monitoring Report. The ideas contained herein necessarily build on a number of projects in which the author has been involved over the past decade, some of which have been supported by UNESCO and other agencies through the International Literacy Institute (see Wagner/ILI, 1998; UNESCO/ILI 1999, 2002a,b). Other work in this area was supported by Unicef, World Bank and the UN Statistical Office; and most recently by the U.S. Department of Education and the Spencer Foundation. Thus, a number of the ideas, concepts and examples presented here have appeared in various ways in previous reports by the author. Nonetheless, all the ideas are those of the author and do not necessarily represent those of UNESCO or any other organization.

List of Abbreviations and Acronyms

BLS	Basic Literacy Skills
EFA	Education For All
IALS	International Adult Literacy Survey
ICT	Information and Communications Technology
IEA	International Educational Achievement
LAMP	Literacy Assessment and Monitoring Programme
LAP	Literacy Assessment Project
LDC	Less Developed Country
NFE	Non-Formal Education
NGO	Non-Governmental Organization
OECD	Organisation for Economic Co-operation and Development
UN	United Nations
UNESCO	United Nations Educational, Scientific and Cultural Organization
UIS	UNESCO Institute for Statistics

TABLE OF CONTENTS

EXECUTIVE SUMMARY

1. INTRODUCTION	5
2. INTERNATIONAL STATISTICS ON LITERACY.....	5
3. POLICY ISSUES IN LITERACY MEASUREMENT.....	7
3.1 LEARNING ACHIEVEMENT AND INFORMATION STAKEHOLDERS	7
3.2 INTERNATIONAL COMPARABILITY OF DATA	9
3.3 LANGUAGE POLICY AND MULTILINGUALISM	10
3.4 MOVING FROM MONITORING TO DIRECT ASSESSMENT.....	10
3.4.1 <i>The end of a dichotomy</i>	10
3.4.2 <i>Large-scale literacy assessments: IALS and successor methods</i>	11
3.4.3 <i>Challenges to the IALS model</i>	12
3.4.4 <i>Household surveys and programme evaluation models</i>	13
4. CHOICE OF DESIGN FOR LITERACY DATA COLLECTION.....	13
4.1 CHOICES IN SURVEY DESIGN.....	13
4.2 A TAILORED APPROACH TO SURVEY DESIGNS: THE SQC MODEL.....	14
4.3 FURTHER COST CONSIDERATIONS IN THE SQC MODEL	17
4.4 LIMITATIONS OF THE SQC MODEL	17
5. CONCLUSIONS	18
6. REFERENCES	21
1. ANNEX: DEFINING AND MEASURING BASIC LITERACY SKILLS.....	25
1.1 DEFINITIONS OF BASIC LITERACY SKILLS	25
1.2 OPERATIONALLY DEFINING LITERACY (READING AND WRITING).....	25
1.3 OPERATIONALLY DEFINING NUMERACY	27
1.4 A NOTE ON 'LIFE SKILLS'	28
2. ANNEX: SOME PROCEDURAL ISSUES IN MEASURING BLS	30
2.1 LEVELS OF SKILLS	30
2.2 READING AND WRITING ASSESSMENT, WITH LEVELS AND EXAMPLES	31
2.3 NUMERACY ASSESSMENT, WITH LEVELS AND EXAMPLES	33
3. ANNEX: CAPACITY BUILDING FOR MONITORING AND MEASUREMENT	35
3.1 HUMAN RESOURCES	35
3.2 TEST DEVELOPMENT, PRETESTING AND PILOT STUDIES	35
3.3 ENSURING QUALITY: DATA COLLECTION, PROCESSING AND ANALYSIS	36

EXECUTIVE SUMMARY

The World Conference on Education for All (WCEFA), held in Jomtien in 1990, found that the measurement of learning achievement was critical to judging the quality of education programmes around the world (UNESCO, 1990). In the decade that followed WCEFA, a number of major initiatives began to reinforce not only capacity building in learning assessment, but also the reconceptualization of what assessment ought to be in various education domains. Some of these formed the basis for a renewed call for increased work on adult literacy, which was part of the World Education Forum, held in Dakar in 2000.

With a UN Literacy Decade launched in 2003, there is even more emphasis being put on ways to ‘keep track’ of rates or levels of adult literacy around the globe. The present paper presents an overview of the main approaches that have been taking to measure adult literacy within and across countries. Since other reviews exist, the primary focus of the present paper is to review the differing purposes of different models of assessment in light of their intended policy goals. There are a number of different approaches to data collection and analysis, each with costs and benefits – where costs are not only fiscal, but also include human resources, time, political capital, and where benefits, in a parallel fashion, may be seen in terms of national rankings or in improving instruction or in teacher training.

A particular focus of the present review is to compare and contrast three models of literacy assessment: (1) the ‘traditional’ model based largely on census enumeration, which has been used over recent decades to collect national and regional rates of illiteracy; (2) the large-scale survey techniques employed with the International Adult Literacy Survey (IALS) and similar efforts; and (3) an intermediate type of assessment that borrows from the previous two models, and attempts to ‘tailor’ the size and complexity of the survey methodology to the policy questions needing answers (called the SQC model).

Various annexes are also included that provide more specific information on definitional, procedural issues, as well as on building human resources capacity for literacy assessments.

The present paper suggests that there is no single model or set of methods that are universally appropriate to monitoring and measuring in adult literacy around the world. On the other hand, there are serious choices to be made depending on ones goals and means for reaching them. In an era with restricted financing for research and evaluation, it is argued, there is an ever-increasing need to ascertain and understand measurement choices and apply them to appropriately to the situations at hand.

1. INTRODUCTION

The World Conference on Education for All (WCEFA), held in Jomtien in 1990, found that the measurement of learning achievement was critical to judging the quality of education programmes around the world (UNESCO, 1990). This recommendation followed on decades of evaluation studies that often used ‘participation’, ‘funds spent,’ or ‘contact hours’ as proxy variables for determining the quality of a nation’s efforts to improve education. With the arrival of WCEFA, it was agreed that what the learner actually learned should be at the center of the educational enterprise.

In the 15 years that have followed WCEFA, a number of major initiatives began to reinforce not only capacity building in learning assessment, but also the reconceptualization of what assessment ought to be in various education domains. Most of this effort, including fiscal support, went into the formal schooling sector. However, soon after Jomtien efforts began to address assessment issues in non-formal education (NFE) and adult literacy in developing countries (Wagner, 1990, Ziegahn, 1992), and major surveys were undertaken in the United States (Kirsch et al., 1993) and internationally (OECD/Statistics Canada, 1995). Some of these formed the basis for a renewed call for increased work on adult literacy, which was part of the World Education Forum, held in Dakar in 2000.

Thus, at the time of the UN Literacy Decade in 2003, a variety of approaches or models of ‘keeping track’ of rates or levels of adult literacy have been put forward, and some have been implemented in various contexts. The present paper contains an overview of the main approaches that have been taking to measure adult literacy within and across countries. Since other reviews exist, the primary focus of the present paper is to review the differing purposes of different models of assessment in light of their intended or presumed policy goals. In short, as with all measurement tools in the social sciences, there can be quite different approaches to data collection and analysis, each with costs and benefits – where costs are not only fiscal, but also include human resources, time, political capital, and where benefits, in a parallel fashion, may be in seen in terms of national rankings or in improving instruction or in teacher training.

In sum, there is no magic bullet in monitoring and measuring in adult literacy, but, as will be argued here, there are serious choices to be made depending on ones goals and means for reaching them.

2. INTERNATIONAL STATISTICS ON LITERACY

In the social sciences, the gathering of statistics on anything – literacy or otherwise – can imply a rather varied set of approaches to data collection. Historically speaking, and in order to provide worldwide comparisons, the international development community has relied over decades almost entirely on data provided to UNESCO by its member countries (UNESCO, 1983). These countries, in turn, typically rely on a *national population census model*, which most often determine literacy ability by self-assessment questionnaires and/or by means of a proxy

variable utilizing the number of years of primary schooling (i.e., 5 or 6 or 8 years of primary schooling equals a 'literate' person). Such data are then collected and collated by UNESCO to create adult literacy rates, typically the number of persons over 15 years of age in the population who are designated as 'literate' divided by the total number of persons in that same age category (and then often broken down by age range, gender, urban-rural residency, and so forth). Considering these same statistics over decades has provided a statistically robust way *monitoring* literacy levels on a national, regional and international basis.

Note here the use of *monitoring*, which from the Latin word *monere*, means "to warn" or to observe on a situation. This implies gathering sufficient, but minimal levels of information needed in order to judge if there is a problem to be warned about. And this is precisely the use of international statistics on literacy to date, as much of the EFA Monitoring effort, of which this paper appears within. Monitoring, thus, implies a 'lite' version of data collection, sometimes using not only proxy variables (such as school statistics which may be only related to, but not directly indicative, of individual literacy levels), but also, especially in the case of self- or other-assessments (where a village leader may 'say' that so-and-so person 'is literate') leave much in doubt as to the actual status of literacy within individuals or groups of individuals.

Because of doubts about the reliability of such data collection, considerable concern has been expressed about the credibility of literacy statistics. In 1986, for example, UNESCO and the UN Statistics Office (UNSO) held a joint seminar in Paris to discuss the use of household surveys to improve the collection adult literacy statistics; a technical report which was the basis of this seminar was later published (UNSO, 1989); these discussions were followed by similar ones over the next decade and a half (see, for example, ILI/UNESCO, 1999, 2001, 2002a,b). The concern, as noted above from the 1990 WCEFA, was whether actual skill learning had taken place; the traditional method of monitoring international literacy levels was clearly very limited since few countries bothered to actually *measure* individual skills in a large or broad enough population sample to assure that literacy levels were valid and reliable (from the Latin, *mensura*, the connotation is one of a system for precise quantitative measurement, as in a tape measure).

Over nearly two decades, there has been a movement to add greater statistical strength to the monitoring of adult literacy, and this has meant a commensurate movement toward direct measurement or assessment of literacy skill. The two terms – monitoring and measuring – are therefore complementary: you can monitor with either good or poor measures, but it is better to have the best measures possible within available resources. This is much easier said than done. In fact, even as many specialists now agree that exclusive reliance on traditional indirect measures of literacy may be flawed, there is renewed discussion of the utility of proxy measures (Murray, 1997, Desjardins & Murray, 2004), since they are may be sufficient and cost less. Indeed, cost effectiveness is an issue that must be kept clearly in mind, as will be discussed below, since resources are always limited, and seeking to collect too much data may be as counterproductive as collecting too little – as will be discussed below.

In sum, the gathering of international statistics on literacy poses a variety of problems which have been matters of debate among specialists. With the advent of greater need for direct measurement, and increased technical capacity for such measurement in developing countries, a

clarification of issues seems to be taking place, as described below.

3. POLICY ISSUES IN LITERACY MEASUREMENT

3.1 Learning achievement and information stakeholders

At WCEFA in 1990, a number of educational targets relating 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. Unfortunately, as noted in the *Literacy and Adult Education* thematic paper for the 2000 Dakar EFA meeting (Wagner, 2000), the Jomtien emphasis on learning achievement had not succeeded in becoming a major part of literacy work in the decade between 1990 and 2000. By 2000, it was agreed that the field of adult literacy requires both a greater focus on the quality of literacy services and better ways to measure literacy achievement. In the Dakar Framework for Action (UNESCO, 2000), these approaches were reinforced by three of the six main stated Dakar EFA goals, namely:

(iii) ensuring that the learning needs of all young people and adults are met through equitable access to appropriate learning and life skills programmes;

(iv) achieving a 50 per cent improvement in levels of adult literacy by 2015, especially for women, and equitable access to basic and continuing education for all adults;

(vi) improving all aspects of the quality of education and ensuring excellence of all so that recognized and measurable learning outcomes are achieved by all, especially in literacy, numeracy, and essential life skills.

Moreover, in the declaration of the UN Literacy Decade (April 2003), monitoring and evaluation processes are expected to include at least three indicators of literacy progress, namely the:

- change in the absolute numbers and in the percentages of the literate population;
- relative contribution of formal and non-formal education to attaining literacy;
- impact of literacy on the quality of people's lives.

To achieve these indicators will require, at a minimum, a way of improved measurement tools over the traditional methods described in the Section 2. With improved assessment methodologies and greater availability of new technologies for collection and analysis of data, it should be possible, more than ever before, to increase access to credible data on learning achievement.

As in all areas of education, and perhaps more salient than in some, there are a number of diverse and even contentious stakeholders in the literacy field. There are, as noted up to this point in the discussion, international and national agencies that seek to better monitor trends in

literacy, so as to make improved decisions about future investments. In addition, there are an even wider variety of local (or international) non-governmental organizations (NGOs), programme directors, practitioners who have a need not only for government political support, but also to improve their ability to function competently and to demonstrate to themselves and others their capabilities. And, last but not least, there are learners, and their organizations, that increasingly call for improvements in the quality of their programmes, and who wish to know what the likely impacts of participation in such adult literacy efforts.

Thus, it seems reasonably clear that neither NGOs nor adult learners living in Botswana will care very much about whether Botswana's rate of literacy is superior in a given year to that of Peru. Similarly, it is not clear that UNESCO will have the resources, under traditional literacy data gathering, to assist Botswana in reaching marginalized ethnic groups living on its borders. The point here is simply that not all statistical methods will likely please all stakeholders equally or at all. Conversely, there are stakeholder needs that can and should be taken into account that may go well beyond the typical national monitoring schemes currently in place.

Therefore, the problem is not – as some have occasionally said – to be only one of sufficient technical expertise or human resources in the field. Rather, there are important choices to be made about which information stakeholders will be attended to in new measurement efforts. To be concrete, let us consider the following types of non-trivial stakeholder questions:

- At the national level. How can we better judge the current status of literacy levels among out-of-school children, youth and adults, irrespective of former school attendance?
- At the programme level. How can we identify skill gaps and needs that may serve as better targets for interventions across diverse ethnic and linguistic groups;
- At the learner level. What am I going to get out of participation in XYZ programme, especially in light of the work needs that I have in my household?

Naturally, policy goals at whatever level, will vary across countries, as well as by gender, ethnic group and region (and more) within countries. Thus, countries concerned about overcoming inequalities caused by geographical disparities may want to collect more information about type of housing or community, or distance from the nearest school, in order to be able to identify those living in remote or in inaccessible areas. Countries concerned about inequalities by gender, ethnicity, or language groups will need relevant variables that capture group membership in this regard. Finally, countries interested in relating the assessment results to specific programmes or educational experiences will have to include relevant questions concerning programme curricula.

Improved literacy measurement and data collection can provide better answers to a variety of stakeholders, but not every method or model can all needs of all literacy stakeholders. Priorities will need to be discussed and set, decisions taken, and some options foregone in order to achieve the possible. The point here is not to bemoan the difficulties of not being able to achieve each and every goal; rather, the point is to reinforce the need to provide a set of options

that may be chosen among in order to meet stakeholder interests. Irrespective of such interests is the requirement to improve measurement tools, a topic to which we now turn.

3.2 *International comparability of data*

The comparability of data is a major concern for policymakers and planning agencies. If definitions and classifications vary, then it can be difficult if not impossible to compare data collected through different surveys. Comparability and stability are necessarily the hallmarks of the UN data collection, including EFA Monitoring and the UIS. Nonetheless, if comparability becomes the primary goal, while less attention is paid to the (local and cultural) validity of the definitions and classifications of literacy, then the data may become less meaningful and potentially less applicable at the ground level. This is a natural and essential tension between ‘emic’ and ‘etic’ approaches to literacy measurement (Wagner, 2004b).

International and national needs, definitions, and research strategies may or may not come into conflict over the issue of comparability, depending on the particular problem addressed. For example, as mentioned above, UNESCO solicits literacy data worldwide, where literacy has been measured in terms of the number of “literate” and “illiterate.” For most countries, this dichotomous type of classification presents few practical (technical) problems and is relatively inexpensive to gather (as part of nationwide censuses), while providing international agencies with a cross-national and time-series framework for analyzing literacy by geographic or economic world regions.

However, educational planners may want to know about the effects of the completion of primary or secondary schooling (how much was learned in a particular area of study, or in a particular literacy campaign the levels of literacy attained. In these cases, a simple dichotomy is too a blunt statistical instrument; skill scores or levels are clearly required for learning achievement to be adequately measured. Furthermore, precise data are needed as to which languages and scripts are used in each region and by ethnic group, in addition variation by age and gender. The collection of such data has largely been ignored by most national and international agencies to date (with some notable exceptions).

The impact on educational policy of such comparative studies is due at least in part to the realization among national policy makers that their country (or population segments therein) may be far lower in certain areas of learning achievement than would have been predicted, say, by participation in school. The 1994 U.S. National Adult Literacy Survey found, for example, that 50% of minority group adults in America with a 10th grade education could read at only a 5th grade level. Issues of equity in the U.S. continue to dominate educational policy, as minority youth are, on average, about four years behind achievement levels of same-age White children, even up through college years (a Black college graduate in the U.S. reads about the level of a White high school graduate). In a World Bank national household survey in Bangladesh, Greaney found that five years of primary schooling resulted in only a first grade equivalent of learning achievement, and that three years of schooling had approximately zero value in terms of learning achievement (Greaney et al., 1998). This study may have an important impact on the

kinds of investments that Bangladesh makes in the area of basic and non-formal education in the future.

3.3 *Language policy and multilingualism*

Most countries have formulated an explicit language policy – one which typically states which language or languages have official status. Often the decision on national or official language(s) is based on such factors as major or dominant linguistic groups, colonial or post-colonial history, and the importance of a given language to the interests of economic development. Official languages are also those most commonly used in primary school, although there may be differences between languages used in beginning schooling and those used later on. Further, there may be important differences between language policy in primary schooling and that of literacy or non-formal education (NFE) for adults. Which languages ‘count’ in literacy statistics may also be controversial. As noted above, when UNESCO collates literacy statistics, it typically depends on government census bureaus to provide the number of literates and illiterates, as well as age and gender differences. In an age of serious ethnic tensions, it may be very important to know in which languages populations can read and write.

While there is usually general agreement that all official language(s) ought to be assessed in a national literacy survey (e.g., English in the United States; English and French in Canada, and German, French, Italian and Romanch in Switzerland), there may be disagreement over the assessment of literacy in non-official or semi-official languages, where these have a recognized and functional orthography (e.g., Athabaskan in Canada, or Telugu in India). In many countries, there exist a multitude of local languages which have varying relationships and status with respect to the official language(s). How these languages and literacies may be included in a national literacy survey can be a matter of serious debate. For example, in certain predominantly Muslim countries in sub-Saharan Africa (e.g., Senegal or Ghana), the official language of literacy might be French or English, while Arabic -- which is taught in Islamic schools and used by a sizable population for certain everyday and religious tasks -- is usually excluded from official literacy censuses. Similarly, literacy in Chinese, Spanish, Cherokee and other written languages have generally been ignored in literacy assessments in the United States.

3.4 *Moving from monitoring to direct assessment*

3.4.1 The end of a dichotomy

We have noted already that literacy, as with other educational subfields, has at its disposal an increased array of measurement tools. Monitoring tools using self-assessment and proxy variables have come under intense criticism over the past decades, and this has given rise to a number of major efforts to conduct large-scale national and international surveys of literacy,

where direct skill assessment is the focus of effort. There is little doubt among both experts and policy makers of the relevance and increased precision afforded by direct measure of skill, as contrasted with the uncertainty and guesswork of the past.

There is little doubt, as well, that the use of dichotomous variables in literacy work have had a deleterious effect on the field in a number of ways: from the initial equating of illiteracy with uncivilized, to the continued use of 'literate' versus 'illiterate' in census data in many countries today (Wagner, 1990, 2001). Indeed, one can nearly argue that in today's world we are all illiterate in one way or another, with the merger of such terms as reading, health literacy and technological illiteracy; or that we are all literate, in the sense that very few people in today's world have little or know knowledge of the purposes and nature of literacy in at least one script. In other words, there are a great deal more shades of gray than black and white in contemporary discussions of literacy and its measurement. This conclusion has the positive consequence of demonstrating that, like formal schooling, adult literacy programming deserves a great deal more investment than it is currently getting (for a current discussion on sub-Saharan Africa, see Lauglo, 2001). With very little doubt, the next decade will see end of present use the literacy-illiteracy dichotomy in discussion of literacy assessment and policy.

Similarly, there is little doubt now that there must be greater resources invested in the direct measurement of BLS achievement. Yet, what type of investments should be made? We begin with a brief review of one of the best-known attempts to engage in direct measurement.

3.4.2 Large-scale literacy assessments: IALS and successor methods

The International Adult Literacy Survey (IALS) has become well-known over the past decade for its emphasis on direct measurement, and its approach to providing comparative data across countries (mainly in industrialized countries; see OECD/Statistics Canada, 1995, 1997, 2000). The IALS methodology is based on a number of national predecessors, such as the 1993 U.S. National Adult Literacy Survey, or NALS; Kirsch et al., 1993), which invested significant resources in improving the technical and psychometric properties of literacy assessment instruments, using a variety of techniques, including methods for expanding the range of items used in a survey format, including Item Response Theory (IRT). The IALS, and its predecessors, utilized a five-level categorization method for literacy, along three different scales (prose literacy, document literacy, and quantitative literacy (or numeracy). These survey scales are not without critics (e.g., Reder, 2000, on the NALS, on the colinearity of the three scales); and there are criticisms, on, for example, the degree of international comparability (e.g., Kalton et al., 1998, on population sampling difference across IALS countries) or on item comparability (Levine, 1998).

Since the launch of the UN Literacy Decade, the UNESCO Institute for Statistics (UIS) has launched an initiative called LAMP (Literacy Assessment and Monitoring Programme), where they are planning to build on some of the tools developed through the IALS, but refocused on literacy assessment in developing countries (Terry, 2003). While the effort has only recently begun, it holds the promise of being able to employ some of the best technical tools in literacy

assessment, and adapt them for use in poor countries. Cautionary remarks about the cost-benefit trade-offs in poor countries have been made by a number of literacy specialists (e.g., Street, 1996; ILI/UNESCO, 1999). Among the issues invoked in such remarks are the low degree of transparency of the data when collected using the IRT methodology, the expensive process of creating methods of cross-cultural comparability, the long time taken to complete a given study, and the overall cost of highly sophisticated methods in poor countries.

3.4.3 Challenges to the IALS model

There are numerous challenges to any method of testing. These range from disputes about theoretical and operational definitions to the type of statistical tests employed to how to analyze datasets. Given space limitations, it is useful to simply note some areas that have been particularly debated in studies involving the IALS assessment methodology. These include: (a) scales of literacy achievement (from dichotomous, to 5 levels, to many levels); (b) determination of when a 'level' is achieved (e.g., in IALS, is it adequate to say that a level is achieved if and only if 80% of the items in a level are completed successfully; see Levine, 1998); (c) what is included in the operational definition of literacy; (d) effectiveness of the use of proxy measures (Lavy et al., 1995); Murray, 1998; and others); and (e) determination of which populations are or are not included in the population survey (e.g., are ethnic minorities, and their languages/literacies excluded?).

The issue of population sampling also poses a set of changes in the IALS model. For example, resources might be invested in a more selective fashion (directing more funds to preschool and primary schools, or to specific groups of adults), so that some individuals -- those with a greater chance of success -- would have access to the possibility to become literate or more literate. Indeed, recent evidence on volunteer literacy efforts in the United States suggests that the least literate portion of the population is remarkably resistant to literacy training, often exhibiting much higher rates of programme attrition and lower learning achievement levels (Venezky, 1992). Research in LDCs in this domain would be very illuminating, and might result in some new policy decision-making.

International surveys like the IALS have also been criticized for being too expensive and too complicated to be 'owned' (that is accepted for endogenous and locally sustainable use) by national and local agencies. While comparative surveys have often received considerable media and policy attention, and have led at times to significant national educational policy shifts, the cost is high relative to local LDC budgets and/or opportunity costs. National or local household surveys can also have a similar policy impact (as in the U.S. NALS), but this result necessitates a serious and credible study, followed by concrete efforts to publicize results, something often difficult to achieve without the 'credibility' of external agencies and high-priced consultants.

The costs of large-scale assessment studies are quite variable. Estimates of the total cost of the IALS survey(s) run as high as tens of millions of U.S. dollars, while the costs of the UNSO (1989) national literacy survey in Zimbabwe may be estimated at about US\$100 thousand

in current dollars. Costs clearly depend on the number of countries included, the degree of external expertise required, the complexity of study design, and degree of collection and analyses undertaken. Clear trade-offs are available in costing processes, from limiting sample sizes to the length of tests created to the degree of trained personnel required. Nonetheless, there is little exactitude in current cost estimation due to the paucity of available studies with cost figures.

3.4.4 Household surveys and programme evaluation models

Household surveys have been used for decades, often employing randomized samples to gather specific types of information on ‘target groups’ within countries or regions in countries, and stratified along certain desired demographic parameters. In the literacy field, one of the first household surveys was undertaken in Zimbabwe, and referenced earlier (UNSO, 1989), with numerous others to follow (e.g., in Morocco, Lavy et al, 1996; in Bangladesh, Greaney et al., 1999; in Botswana, Commeyras & Chilisa, 2001). Further, in a multiyear effort just preceding the 2000 Dakar EFA Forum, the International Literacy Institute (ILI) and UNESCO undertook a series of workshops and field studies in the Literacy Assessment Project (LAP). LAP took a pro-local approach to surveys, trying to situation data collection more toward meeting local and national needs, and practically based tools that could be understood by laypersons and not-highly-trained adult education specialists. This position was seen to contrast with the IALS model, designed primarily to deliver data to high-level national and international agencies. A number of reports were published in support of LAP approach (see ILI/UNESCO 1999, 2001, 2002a, b; ILI, 2000).

A more common approach at the ground level is that of programme evaluation methods, which have a long history in adult literacy and educational development work. Unfortunately, with a few exceptions (such as the recent and interesting Okech et al. work in Uganda, 2001), most programme evaluation work does not use sufficiently robust methods that would allow for serious skill assessment. This is not surprising in that evaluations are largely sponsored within NGOs which have little in the way of internal technical capacity.

4. CHOICE OF DESIGN FOR LITERACY DATA COLLECTION

4.1 *Choices in survey design*

There are, of course, many ways to collect data on individuals and groups of individuals. Choosing among them is a major challenge which is (or should be) directly related to the policy or programme questions that one seeks to answer. A brief summary of these models is provided as follows.

- *National and international sample surveys.* The IALS model, as described in some detail above, is most often undertaken as part of the work of a national census bureau, with a focus

on proper sampling across demographic parameters, and the inclusion of psychometric tests and analytic techniques. Efforts to make such surveys comparable at an international level are complex and relatively expensive.

- *Household surveys.* These allow considerable flexibility in terms of sample selection and sample reduction (one can limit the sample size through appropriate stratification), and thereby provide more time for actual learning assessment. One limitation is that such surveys are often ‘stand-alone’, and require specialized training as well as relatively skilled test designers. As part of this same approach, *special literacy surveys* may be undertaken with a particular focus on literacy assessment, but neither are part of a census sample (household or otherwise). One example of this approach was that undertaken by the World Bank (Greaney, et al., 1999) in rural Bangladesh, or as part of research projects (Wagner, 1993). A recent derivative of the household survey has been termed the SQC approach, which will be discussed in the next subsection.
- *Post-censal sample.* The LAP India case study cited the NSSO study which was of the post-censal (after a census) type. One advantage is that the same personnel and data structure can be used as in the census, such as in the India case study in the LAP project (ILI/UNESCO, 2002b), there are advantages in being able to compare the data collected from the regular census with the post-censal in-depth study. One limitation is that census personnel often seem to have less technical capability in designing BLS assessment instruments.
- *Programme evaluation.* Evaluations of literacy programmes in developing countries are numerous, of both formative and summative varieties (Bhola, 1990; Carron et al., 1989; Okech, et al, 2001; Nordtveit, 2004). In developing countries especially, such evaluations have only rarely included psychometrically appropriate tests for measuring learning achievement. Typically what is measured are the inputs in human and infrastructural resources, the pedagogical methods employed, and the outcomes in terms of attendance and ‘successful program completion.’ A lack of skill assessment is typically a very serious gap in the current programme evaluation knowledge base.
- *Other issues.* Among other issues is how to meet high-stakes versus low-stakes concerns on the part of participants in the assessments. In-school tests are often seen as high-stakes, but this is not the purpose of the current literacy assessment exercise – yet participants often view testing as a simple extension of schooling. Thus, care needs to be taken, or, as in some of the case study examples, one may find data falsification on the part of either learners or instructors or both. Finally, some procedures for post-test adjustments for sampling biases may be undertaken, though these may or may not be necessary, depending on the sample and testing methodology chosen.

4.2 *A tailored approach to survey designs: The SQC model*

As alluded to above, it is clear that international and local needs may not be one and the same, with ‘national’ needs falling somewhere in-between. Countries and cultures are diverse,

each with a multiplicity of groups that vary along ethnic, linguistic, social class, economic and other dimensions. Each country has its own special history of sociopolitical development, and its own experiences with formal schooling and broader educational development. The international policy community has its interests as well, mostly in trying to guide national decision-making from indices of where nations “fall” on some scale of economic productivity or worker efficiency – hence the ‘horserace’ concept in international comparisons.

The improvement of literacy assessment in comparative context may affect local, national and international interests in contrasting ways. National interests and “internal” considerations (involving, for example population/ethnic diversity) may be seen as nettlesome problems or simply constraints by planners concerned with international comparison. On the other hand, national considerations about population diversity, linguistic variations, and even orthographic diversity (such as unusual features of a script) may be seen as having to be sacrificed on the methodological altar in order to achieve a larger basis for international comparison. For these and other reasons, there is ample logic for local programmes and national level policy makers to hesitate in sacrificing local interests for those with an interest in regional or international comparisons.

More specifically, the level of resource investment in empirical data gathering in IEA-like studies is, for many developing countries, far greater than that made heretofore. Thus, there may be opportunities to create a research infrastructure through the carrying out of international comparisons, not so different in kind from that achieved by anthropologists working with diverse groups in developing countries. Perhaps most importantly, comparative studies can, if properly designed, help to achieve a greater understanding of cultural values and attitudes towards learning and literacy. Such an approach would enable the problem of diversity to be turned into opportunity of studying varieties of learning in context.

The above analysis has led to the importance of seeking alternatives to the technically complex and expensive IALS model on the one hand, and simple non-psychometric programme evaluation on the other. What might be called a compromise model is, in reality, an approach that is tailored to the nature and size of the policy and programme questions at hand – what has been termed the Smaller/Quicker/Cheaper (or SQC) model (Wagner, 2004a).

a. Smaller. Assessment methods do not need to be major entrepreneurial enterprises, but rather just robust enough to answer key policy questions at the national and local levels. Indeed the focus on ‘size’ needs to be tailored, as with all social science methodology, to the precise set of questions to be answered. The term ‘smaller’ generally has two main meanings: first, the number of countries included in such studies may be only one, and in the case of large countries (e.g. India), may be at the state or sub-state level. Second, whatever the population group studied, the population of human assessed, as well as the number of items utilized in assessment instruments need only be ‘just large enough’ to answer the relevant questions. Of course, in some cases, this may be easier said than done – it is not always possible to gauge such matters with great precision.

Nonetheless, many international comparative studies, such as the IALS, seem to run counter to this perspective. That is, the lowest common denominator is often to ‘go large’ so as to be sure to have enough data, especially if some data is ‘tainted’ for one reason or another. This rationale is reasonable and plausible, but would be much less likely to occur if the study was done on a smaller scale. Just as importantly, large-scale international studies often cannot literally afford to focus on those ‘most in need’, such as minority ethnic groups, displaced persons or out-of-school girls. Here, the emphasis on large and comparative assessments may work against the stated policies of international agencies such as the World Bank and UNESCO. If strong empirical science – as exemplified by numerous countries, large datasets, and heavy inferential statistics – works against policy goals, then these approaches need to be given some reconsideration. In other words, ‘big’ can be ‘bad’, from a development perspective.

b. Quicker. Literacy assessments need to be completed in ‘real time’, or thereabouts, so that results can affect policy and spending in the ‘lifetime’ of current ministerial appointments and/or programmatic needs. Studies that take 3-5 years to generate results, even if analytically robust, nonetheless fail to meet the test of timeliness. The IALS and International Educational Achievement (IEA) studies, just to name a two examples, usually take years to carry out, analyze and publish. The time-span needed to make local decisions that affect budgetary policy and programmatic change is much shorter, usually at most 18 months. Reconciling these conflicting temporal needs would not be so difficult if the studies were smaller in the first instance. Large-scale studies necessitate more time; but are large-scale studies necessitated themselves?

c. Cheaper. Funding is a driving force in all development work, and is one of the reasons why large-scale assessments have received the large majority of funding for literacy assessment. It is much easier to be on the ‘radar screen’ of the World Bank if one has a large-scale study, with the potential for a large policy report or reports. It is also an easier way to handle development agency budgets. But seen from a national or local perspective, things may be quite different. LDCs may not be able to afford either the fiscal or human resources costs of deep involvement in highly technical assessment exercises over multiple years. Further, the higher the cost, the more difficult it is to get to an initial ‘yes’ to participate in such a national exercise, and the more difficult to gather time series data to follow policy decisions.

Cheaper also means here the possibility of using local (rather than external) experts and consultants. This can be achieved if the assessments are not constrained to use outside agencies in the industrialized countries to process complex data sets. By choosing simpler psychometric methods, one can make data and their analyses more ‘transparent’ to local human resources. Assessments necessarily comprise a variety of expertise. These include the policy makers, psychometricians (test makers), survey and interview specialists (enumerators), and data analysts, as well as learners and instructors for preparatory pilot testing. Capacity building -- the training of sufficient expert staff -- has been a major impediment in past large-scale surveys. How such resources can be put into place in the future will be a major question to address, but it will be much easier in local and national contexts than on an international level.

4.3 *Further cost considerations in the SQC model*

In addition to the general cost-reduction features of SQC, it must be said that each design consideration in a literacy assessment has its specific associated costs. Clearly, for use in developing countries, the intent here is that such literacy assessment can be low-cost relative to other, “higher cost” approaches – hence the term ‘cost-effective’ that has been used throughout this document. However, it needs to be said at the outset that there are no absolute certainties with respect to cost, though there are some trade-offs that can be made in terms of cost and quality. The following are specific cost considerations with respect to the SQC model:

- 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 especially those which include international comparisons – can drive costs upward.
- Lower levels of statistical reliability than sometimes used in censuses may be acceptable in literacy 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.

4.4 *Limitations of the SQC model*

How can both comparability and context sensitivity be appropriately balanced in literacy and basic skills assessments? Indeed, how can EFA monitoring, or UN statistics collection be maintained as stable and reliable when localized approaches are chosen over international comparability? The answer would seem to lay, as it should, in compromise. At present, much if not most “comparative” data from LDCs is inherently flawed by failure to use direct assessments. The SQC model would rectify this situation by encourage all countries to choose some version of an SQC survey to enhance the credibility (validity and reliability) of data collected at the national level. These data could further be compared to similar data from other countries. However, what is *not* being required in the SQC model (but which I required in the IALS model) is the strict item by item comparability. Under SQC the advantages of size, speed and cost – and what follows from such agility – are thought to outweigh the importance of over-exactitude in cross-national comparison. The SQC is an improvement on the present system, but seeks to limit its ‘overhead’ in other expense categories, including opportunity costs, of large-scale studies.

One should not confuse the SQC acronym with the word ‘simple.’ There is no question that SQC methodology requires some of the same skilled capabilities as does larger and more complex studies – this is more a matter of degree of intensity and time of inputs. As with the IALS, SQC implementation of a household survey of out-of-school requires careful planning and often human capacity building. Any survey must be planned and implemented in the field in such a way that the information collected is credible and free of error, while costs, logistical efforts, and time needed are kept at reasonable levels. At the same time, the design of data

collection and reports from the survey must be 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 literacy 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 programmes whose projects may serve out-of-school populations.

In contrast to national and international studies where policy makers are the main stakeholders, programme level stakeholders would include directors and teachers in programmes, and possibly employers, parents; learners may also be considered among stakeholders, as they have a vested interest in the quality of the programme 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. While this is a strength of SQC in terms of quality, eliciting support from a diverse set of stakeholders may also present serious problems which may be seen as limitation of the SQC model.

Finally, as described elsewhere (ILI/UNESCO, 1998), the notion of “shareability” of assessment tools is relevant here. Tools developed for SQC household surveys can and should be “shared” with programme 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 programme evaluations. When shareability is not achieved, this may also be seen as a limitation, since it represents one of the distinct advantages of the SQC model.

5. CONCLUSIONS

Educational policy makers at all levels typically seek to have the best data resources possible in order to make decisions. The problem in the adult literacy arena, as in others, is how to define ‘best.’ Breadth (e.g., broadest demographic sampling) and depth (e.g. most reading items tested) of data collection are ways of thinking about what is best, but so are transparency, “shareability,” and human resources capabilities – all terms that have been discussed above.

The movement toward improved monitoring and measuring literacy may be summarized by a small set of models to choose from that will, nonetheless, have rather complex consequences. These models are summarized again below:

- a. Traditional model. Traditional census or census-like data collection as over the past several decades, usually taking the form of self-assessment or proxy variables (years of schooling). The advantages are ease of data collection, tradition of collecting data in this manner, stability and reliability (i.e. replicable results); disadvantages include lack of validity (it is clear that self-assessment or proxy measurements are statistically correlated with achievement, but can often be misleading in specific or even general ways).

- b. IALS model. This model, developed originally for use in industrialized countries utilizes the most advanced technical and statistical tools in order to achieve statistical comparability at both international and intra-national levels. Advantages include sufficiently robust data collection such that high-powered analyses can be undertaken. The disadvantages include cost, lack of transparency to diverse end-users, and a challenge to validity due to difficulty of creating equivalencies across languages and cultures.
- c. SQC model. This model had its precursors in a number of research and household surveys over the past 15 years or so. The advantages of SQC are inherently the emphasis on smaller, quicker and cheaper ways of gathering and analyzing data (compared to, say, the IALS model). The disadvantages include less investment in international comparability, and more variation from country to country since the model depends necessarily on the different, country-specific priorities.

There is little question that the IALS and other international assessments have put the direct assessment of literacy on the policy map, as has been suggested for many years (see Wagner, 1990). Yet the successes of IALS have also pointed to some of its limitations both in design and in promise. With a focus on the poorest of the poor in the UN Literacy Decade, an argument has been made here for considering each measurement model on its own merits or disadvantages.

However, if the goal is to improve literacy in the poorest LDCs, a major goal of the UN Literacy decade, the SQC approach may allow greater effectiveness in reaching the unreached and un- (-under) schooled (gender, minorities). Further, the SQC model can foster greater impact of data on policy (robust) than is currently available; have a more timely impact of data on policy (practical); and provide for a more sustainable future of the knowledge base. Much needs to be done to obtain a full vetting of alternative approaches that do not seek to dismantle what has been learned already (and will continue to be learned) from international comparative studies. Yet, a unique emphasis on either the traditional or the IALS approach will miss some important opportunities.

Policy makers will inevitably undertake further surveys in order to answer the following kinds of questions: (a) what does a given population 'know' in terms of literacy skills; b) how these literacy skills are used and (c) what policy changes can be made that might influence future levels of basic learning competencies or ways to improve the acquisition of literacy skills. Such efforts will need not only the collection of skills-related (psychometric) data through the testing instruments, but also careful collection of background information that is essential in order to evaluate the determinants (independent variables) associated with observed skills. Further, if policy makers wish to understand how different levels of prior schooling impact actual (measured) knowledge or performance levels in certain skill domains, they will need to make sure they have reliable data about relevant background variables, such as language spoken at home, social, and prior educational experiences. Impact studies will also require baseline and follow-up assessment procedures.

In sum, there are real choices to be made about literacy monitoring and measurement, and there are different models for different purposes. Since resources in adult literacy education have always been seriously limited, making the right choice for data collection is something that will inevitably take on greater saliency throughout the remainder of the UN Literacy Decade and beyond.

6. REFERENCES

Commeyras, M. & Chilisa, B. 2001. Assessing Botswana's first national survey on literacy with Wagner's proposed schema for surveying literacy in the 'Third World'. International Journal of Educational Development, 21, 433-446.

Desjardins, R. & Murray, R.S. (2004. in press). Introduction to special issue on *Literacy Proficiency: Investigations Using Direct Measures Of Skills*. International Journal of Educational Research.

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.

ILI/UNESCO. (1999). (D. Wagner, principal author). Assessing basic learning competencies in youth and adults in developing countries: Analytic survey framework and implementation guidelines. ILI/UNESCO Technical Report. Philadelphia: International Literacy Institute, University of Pennsylvania.

ILI. (2000). (Principal author, D. Wagner). "Literacy and Adult Education." Global Thematic Review prepared for the U.N. World Education Forum, Dakar, Senegal. Paris: UNESCO. (Reprinted in Adult Education and Development, Bonn: IIZ/DVV, in English, French, Spanish).

ILI/UNESCO. (June 2001). LAP First Experts' Meeting Report. Philadelphia: ILI.

ILI/UNESCO. (April 2002a). LAP Second Experts' Meeting Report. Philadelphia: ILI.

ILI/UNESCO. (April 2002b). Analytic Review of Four LAP Country Case Studies. Philadelphia: ILI.

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.

Lauglo, J. (2001). Engaging with adults: The case for increased support of adult basic education in sub-Saharan Africa. Washington: World Bank.

Lavy, V., Spratt, J. & Leboucher, N. (1995). Changing patterns of illiteracy in Morocco: Assessment methods compared. LSMS Paper 115. Washington, DC: The World Bank.

- Levine, K. (1998). Definitional and methodological problems in the cross-national measurement of adult literacy: The case of the IALS. Written Language and Literacy. Vol. 1 (1). pps 41-61.
- 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.
- Murray, T. S., Kirsch, I. S., & Jenkins, L. B. (Eds.) (1998). Adult literacy in OECD countries: Technical report on the first International Adult Literacy Survey. Report NCES 98-053. Washington, D.C.: U.S. Department of Education.
- Nordtveit, B. H. (2004). Managing public-private partnership: Lessons from literacy education in Senegal. Washington, D.C.: World Bank.
- OECD/Statistics Canada. (1995). Literacy, economy and society. Paris: OECD.
- OECD/Statistics Canada. (1997). Literacy skills for the knowledge society: Further results from the International Adult Literacy Survey. Paris: OECD.
- OECD/Statistics Canada. (2000). Literacy in the information age. Paris: OECD.
- Okech, A., Carr-Hill, R., Katahoire, A. R., Kakooza, T., Ndidde, A. N., & Oxenham, J. (2001). Adult literacy programs in Uganda. Washington, D.C.: World Bank.
- Oxenham, J. (2004, in prep). The Quality of Programmes and Policies regarding Literacy and Skills Development. 2005 EFA Monitoring Report. Paris: UNESCO.
- Street, B. V. (1996). Literacy Economy and Society: A critique. Working Papers On Literacy, No. 1. Montreal: Center for Literacy.
- Street, B. V. (1999). The meanings of literacy. In D. A. Wagner, R. L. Venezky, & B. L. Street (Eds.), Literacy: An international handbook. Boulder, CO: Westview Press.
- SubbaRao, I. V. (2002). India: LAP Country Case Study. Working document published at www.literacy.org. Philadelphia: ILI.
- Terryn, B. (2003). The Literacy Assessment and Monitoring Programme (LAMP): Background and Approach. Paper presented at ANLCI Conference in Lyon, France, (November).
- 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 (2000). The Dakar Framework for Action: Education for All: Meeting our Collective Commitments. Adopted by the World Education Forum, Dakar, Senegal, 26 – 28 April 2000. 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.

Venezky, R.L. (1992). Matching literacy and social policy. Technical Report. Philadelphia, PA: National Center on Adult Literacy.

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

Wagner, D. A. (1998). Literacy retention: Comparisons across age, time and culture. In H. Wellman & Scott G. Paris, (Eds.). Global prospects for education: Development, culture and schooling. Washington, D.C.: American Psychological Association. pps. 229-251.

Wagner, D. A. (2000). Literacy and Adult Education. Global Thematic Review prepared for the U.N. World Education Forum, Dakar, Senegal. Paris: UNESCO.

Wagner, D. A. (2001). Conceptual dichotomies and the future of literacy work across cultures. In C. Snow & L. Verhoeven (Ed.). Literacy and motivation: Reading engagement in individuals and groups. NJ: L. Erlbaum.

Wagner, D. A. (2004a, in press). Smaller, Quicker, Cheaper: Alternative Strategies for Literacy Assessment in the UN Literacy Decade. International Journal of Educational Research.

Wagner, D. A. (2004b). Literacy in time and space: Issues, concepts and definitions. T. Nunes & P. Bryant (Eds.). Handbook of Children's Literacy. Pps. 499-511. Lancaster, UK: Kluwer Academic Press.

Wagner, D. A. & Kozma, R. (2003). New technologies for literacy and adult education: A global perspective. Technical report. Philadelphia: International Literacy Institute/University of Pennsylvania. (also UNESCO, in press).

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

Wagner, D. A. et al. (May 2002). Towards Guidelines for the Improvement of Literacy Assessment in Developing Countries: Conceptual Dimensions based on the LAP Project. Philadelphia: ILI/UNESCO.

Ziegahn, L. (1992). National literacy assessments: present practice and alternatives. International Journal of Educational Development. Vol. 12, pp. 223-232.

1. ANNEX: DEFINING AND MEASURING BASIC LITERACY SKILLS

1.1 *Definitions of basic literacy skills*

The EFA initiative has described needed skills in two differing, but overlapping, ways, namely:

- (i) In 1990, at Jomtien, as “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”; and
- (ii) In 2000, at Dakar, as “literacy, numeracy, and essential life skills.”

These ideas have been explored further in a number of varied reports cited above (OECD/Statistics Canada, 1995, 1997, 2000), in a World Bank survey in Bangladesh (Greaney, et al., 1999), as well as in the LAP documents (ILI/UNESCO, 1998, 1999, 2002a,b). While there is now substantial agreement as to what constitute *operationally definable* core basic skills (reading, writing, and calculating), there is *little or no consensus* across policy makers, countries, specialists or research studies – either within or across countries – as to what constitute acceptable and *operationally definable measures* of such competencies as oral expression, problem solving, knowledge, skills, values, and attitudes, and essential life skills. Further, little if any mention is made of “technological literacy,” a term becoming ever more prominent and important in many countries rich and poor (Wagner & Kozma, 2004).

For the purposes of the present discussion, the focus will be maintained on how to measure those basic literacy skills (BLS) or competencies on which there is already substantial agreement and less debate, namely reading, writing, and calculating. This is not to say that the other skills and dimensions are unimportant for the present or future, but their complexity especially in relation to different cultural contexts would necessitate additional conceptual and methodological development that go beyond the scope of this document. For example, additional work will eventually be required in order to develop some set of standards, or range of performance (including minimum levels of competencies) in literacy skills as related to their functionality in a variety of contexts, as well as in such domains as life skills, work skills and income-generating skills, and so forth. Such discussion also necessitates a greater understanding of the uses of literacy in everyday life as well as the roles that the literacy environment plays in supporting (or not supporting) the functional uses of literacy (cf. Oxenham, 2004).

1.2 *Operationally defining literacy (reading and writing)*

Many definitions exist for literacy and there is an on-going debate on the issue. All relate in some way, at their core, to a person's ability to understand, communicate, and use printed text. 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)

“The ability to understand and employ printed information in daily activities, at home, at work and in the community – to achieve one’s goals, and to develop one’s knowledge and potential.” (OECD/Statistics Canada, 2000)

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; see Street, 1999). 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 adults. 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)
- Ability to understand the relationship between text and images
- 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. Thus, the assessment scheme for literacy elaborates separate tools for reading and writing while at the same time improving our understanding of the relationship between reading and writing. Furthermore, prerequisite and basic levels of literacy cannot be defined in the abstract; 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.

1.3 Operationally defining numeracy

In the past, most definitions of literacy have included calculating skills, with occasional reference to concepts of quantity and other mathematical concepts; thus, most numeracy definitions 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 broad notion of literacy assessment.

Numeracy encompasses a broad range of skills, thought processes, background knowledge (formal and /or informal), and certain attitudes and dispositions. Numeracy enables a person to interpret, act upon, and communicate about mathematical information in a wide range of everyday or work-related and other life contexts, and it thus supports effective participation in a wide range of life roles. It should also be understood that literacy and numeracy are often intertwined, and that literacy skills are typically a significant part of effective deployment of numeracy skills (e.g. in math word problems, graphical displays, etc.). Thus, while it is possible to have numeracy and literacy skills that are independent of one another, most surveys (e.g., the 1995 IALS study) show a strong relationship between the two sets of skills.

Overall, numeracy makes possible critical comprehension of quantitative statements about trends and changes in one's community or nation, and in the world. Numeracy is also needed 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 are 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
- Possession of 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 basic skill 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.

1.4 A note on ‘life skills’

As cited above, the 2000 EFA Dakar Forum included the importance of “essential life skills.” Without further specificity, of course, the term “life skills” could cover an extremely broad range of competencies. These have received relatively little attention from educators as part of the formal school curriculum, however in recent years such skills have been 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 achieve their personal goals, function effectively in their social environment, prevent or cope with situations in which their own (physical or psychological) well-being or safety may be at risk, and enhance their quality of life.

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

- Knowledge of nutrition and eating habits (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 dangerous to your health?).
- Knowledge of personal finances. (e.g., How much savings would be required to build or buy a house in your village?)
- Knowledge about civic rights

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.

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

remains a dearth of research on assessment in the area of life skills, any life skills assessment framework should focus primarily on areas that can be better measured (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. Mention should also be made of cognitive skills that go beyond ‘knowledge-based’ life skills, and include such skills as the ability to search for and access information, critical thinking, and problem-solving – areas alluded to in the Jomtien list of basic competencies (UNESCO, 1990).

Thus, even though the EFA initiative calls for the promotion of life skills and other types of competencies and attitudes, these must be considered during implementation of literacy assessments in ways that enable a clear understanding of the kinds of life skills that are considered essential for the country or local environment. At the same time, it may be noted that while the importance of life skills is widely acknowledged, the development of usable assessment instruments remains a substantial task with much work still to be done to operationally define. At present, even in the well-financed research of the richest countries today, there is little agreement as to the kinds of life skills and problem solving skills that all students should acquire.

2. ANNEX: SOME PROCEDURAL ISSUES IN MEASURING BLS

2.1 *Levels of skills*

A key goal in skill assessment is the development of levels of ability (or proficiency) that may be used to understand the range of competencies across individuals and groups. Further, there has been a concern with ‘how much is enough’ – or the ‘threshold’ of skill necessary for individuals or groups in a given society. However, clearly one cannot have a threshold of learned skills without assessment tools. Further, basic skills, when dealing with out-of-school children, youth and adults, should be both understood and assessed as much as possible in the realm of ‘everyday’ or ‘functional’ competencies even though these terms are necessarily sensitive to local cultural meanings. However such terms are operationalized, literacy assessment must take special care not to apply, *de facto*, tests of competency that are derived directly from school-based practices. The sections below provide some examples of how this strategy might be implemented.

As noted earlier, many approaches to assessment have been tried to date – from the simple dichotomy of literate vs. illiterate, to continuous scales which can provide quite specific scores. The approach developed here – derived from LAP model (ILI/UNESCO, 1999, 2002b) – is a model, which draws on 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 programmes and agencies) will want to ascertain “threshold” levels or targets of skills competency (i.e., what is the minimal goal of “what should be learned or known” in basic education).

Thus, in the literacy skills described below, each is partitioned into four criterion-based categories. The choice of four levels (rather than more or fewer levels) has been provided elsewhere (ILI/UNESCO, 1999), and its two major rationales may be summarized as follows. First, the EFA initiative (Jomtien and Dakar) have called for goals on basic skills; without some delineation of basic skill targets or thresholds, policy makers will not be able to either establish goals or know when they have met them. In the model below, the ‘basic’ level of skill is expected to represent that EFA target, though exact scoring for achieving the level is left to empirical work within each country. Second, there is a need for both limiting cost and increasing transparency (these are in fact related). Even though specific test scores will be generated through any assessment scheme, having too many levels delineated will not only increase costs, but also be less well understood by either policy makers or laypersons. Eventually, it is hoped that with sufficient country level experience, a corpus of concrete test items (even if generic) could be produced for use in developing literacy assessment tools. Furthermore, it should be possible to prepare a set of recommendations based on such experiences that would be able to guide future efforts, such as how to interpret issues of classification and misclassification, what types of items are likely to be more reliable across multiple languages within countries, and ways

to determine which items are most appropriate for out-of-school learners in contrast to those that are more academically based.

2.2 *Reading and writing assessment, with levels and examples*

Reading. Literacy tests – the most common of literacy 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 BLS assessment scheme for reading is based on a matrix of reading skills and domains of print. This matrix 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, and applied skills. 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.); 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. 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”. The former set of skills may be seen as most related to 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. 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
- 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. As noted earlier, advanced skills are equivalent to the level of those who have successfully completed a secondary school curriculum or its equivalent.

Writing. The BLC assessment scheme for writing is based on a matrix of writing skills and domains of print, and 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 and materials more commonly found in schools or many urban environments (e.g., pencil, paper, chalk, slate, keyboard), 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. Writing at this level can typically be measured by having the individual write from 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. 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. 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 (applied) 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
- 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 to perform more complex writing activities (e.g., filling in a complex government form).

2.3 Numeracy assessment, with levels and examples

The assessment of numeracy is similarly based on a matrix of numeracy skills and domains, as well as four ability levels: none, prerequisite, basic, and advanced levels. Numeracy skills may be 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 many who 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. 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. Four levels are proposed to describe numeracy competencies, as follows.

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 numbers.

Prerequisite level. The prerequisite level would include individuals who can engage in some mental calculations, possibly even advanced ones, 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 programmes among out-of-school youth and adults, or on other country-specific factors.

Advanced level. A person could be said to be highly numerate who can, with little difficulty or few errors, accomplish numeracy tasks such as various arithmetic functions in formal school-type and everyday settings, locating and using quantitative information in different displays, solving multi-step calculations, and using measuring devices, among others.

3. ANNEX: CAPACITY BUILDING FOR MONITORING AND MEASUREMENT

3.1 *Human resources*

The implementation of any survey 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. 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 literacy 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. 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 expertise from other countries.

3.2 *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 are 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 offer high information value for literacy assessment, the guidelines would include a number of suggestions, such as the following:

- Develop standards of performance, measurement scales, and basic performance levels.
- Use functional, performance-based tasks and tests, rather than academic, school-based ones.
- Strive for contextual relevance rather than national comparability.
- Develop a specific set of task items that can efficiently measure literacy skills, while acknowledging the fact that there is no universal test that has the same meaning for every subgroup.
- Plan for an on-going process of test construction, revising the test, and validation, since the literacy 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 programmes, national and international agencies).

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

3.3 *Ensuring quality: Data collection, processing and analysis*

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. Once a study is underway, regular discussion and feedback among administration planners, design personnel, and data collectors are essential to ensure both the reliability and validity of the data. Even if rigorous pilot testing is conducted, some data processing should take place early in 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.

At least three levels of data analysis should be utilized:

- To establish reliability and validity of the survey items and instruments (the culmination of the instrument construction process)
- To generate descriptive statistics that provide an overview of literacy skill levels, sub-indexed by major background and demographic variables of the survey (e.g. age, language, region)
- 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 programme level. Policy impact requires an understanding of the consumers (users) of reports on the literacy survey. Governmental agencies and programme 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.