

## REPORT ON UPPINGHAM SEMINAR NOVEMBER 2003

### MEASURING LITERACY: Meeting in Collision

*Note by Alan Rogers: This report has been produced by Carolyn Petersen of the School of Education, University of Edinburgh. She has been assisted throughout by many of the participants who contributed their material, but it reflects her own taking on the Seminar. We are extremely grateful to her for the care she has taken in the preparation of this report. We are all very conscious how difficult the task has been, for the nature of the Seminar made it difficult for any one member to obtain an overview of the whole. All the participants have had an opportunity to see it in its final form and to indicate any major disagreement with any aspect of this report.*

*Uppingham Seminars wish to thank DFID for their suggestion of this Seminar discussion and for their funding which contributed greatly to the success of the event. Other funding came from Education-Action, UIE and UNESCO; and some participants raised funds from other agencies. To all of these, we are grateful.*

#### **About this report:**

This report is divided into three sections. The first is a short report summary of the main points covered. This is followed by the main body of the report describing the discussions with reference to specific presentations given and seminar papers provided, plus the conclusions. Finally, an appendix has been included to give a more detailed account, in order to do justice to the process, presentations and papers provided for the seminar.

#### **Acknowledgements:**

I would like to thank all the participants, who ensured it was an enjoyable and interesting seminar, and particularly Alan Rogers and Jean Baker for their organising skills and welcome. I wish to thank all those who contributed to the report, not least those who diligently took notes and made them available to me after the seminar, including Lindsay Howard, Constant Leung, Jane Freeland and Kara Jackson; and those who contributed with comments, references and articles during the writing up process including Bénédicte Terryn, Vincent Greaney, David Barton, Clinton Robinson and Constant Leung. I would especially like to thank the people who reported back verbally and in note form on small group sessions that I did not attend, including Jane Freeland on the important multi-lingual literacies group. I would also like to thank Barbara Trudell of the University of Edinburgh for her input during the process of clarifying some of the ideas presented here, and Sam Bridgewater during the editing stage. I have endeavoured to represent the full range of views as far as possible within the constraints of the format allowed. Nevertheless, the final report is my own.

Carolyn Petersen  
Centre of African Studies  
The University of Edinburgh  
C.J.Petersen@sms.ed.ac.uk

## MEASURING LITERACY: SUMMARY REPORT

The Uppingham Seminars are a series of conferences held at irregular intervals and organised by Alan Rogers together with a Steering Committee that includes Anna Robinson-Pant and Brian Street. This year's seminar was attended by 32 people from a range of international agencies (including the World Bank, UNESCO, and DFID), research and practice in the field of literacy and numeracy in countries from the 'North' (UK, USA, Ireland) and the 'South' (including Kenya, Uganda, Botswana, India, Nepal, and the Philippines). Discussion papers<sup>1</sup> were sent out beforehand and short presentations given during the seminar.

### Aims:

1. The aim of the seminar was to explore common ground between ethnographic methods and more conventional measurement methods for assessing<sup>2</sup> literacy and numeracy; and to seek out and document ways that ethnographic methods can contribute to the assessment of literacy.
2. The discussions aimed to examine the consequences of a 'multiple literacies' view, including multi-lingual literacies, for the assessment of literacy.

### Context and Process of seminar:

Large-scale surveys assessing literacy and numeracy have been criticised from both the technical point of view and by ethnographers and qualitative social scientists, especially on the quality, validity and comparability of the data (Hamilton and Barton 2000; Street 1996). Nevertheless the pressure is ever greater to produce comparable literacy statistics for international comparison. This is as a result of their use as indicators such as for progress on formal education provision in the Education for All initiative and for progress on the wider development targets in the Millennium Development Goals. International discourse has shown ambivalence in the attention paid to the multiple view of literacies. The following two main approaches were represented at the seminar; the large-scale statistical assessment approach of international agencies such as UNESCO based on an *individual* technical view of *literacy*; and ethnographic and qualitative approaches to assessing literacy based on a multiple *social* view of *literacies*.

### Key Questions:

The event tackled the key questions of: 'Why' measure literacy (for what purpose and who for), 'What' should be assessed or measured, and 'How'. The issue of scale was crucial to this; whether assessment is for the purposes of large-scale comparison or accountability; to inform implementation, or to provide benefits at the micro-scale of the learner.

The seminar created a space for discussion and for participants to develop their views in relation to the range of standpoints. There was no fixed agenda, and no structured previously prepared inputs. This resulted in constructive dialogue between the many different points of view. One outcome of the seminar was that participants seemed to have come to a greater

---

<sup>1</sup> Including a background paper on Measuring Literacy (Kell and Rogers), the Literacy Assessment and Monitoring Programme (LAMP, UNESCO) (Terry), Psychometric testing and Item Response Theory (IRT) (Leung), the Equipped for the Future (EFF) approach (Stites), the National Adult Literacy Agency (NALA) (Merrifield), Multi-lingual literacies in Vanuatu (Jackson), Numeracy as a social practice (Baker), the Kenya Participatory Impact Monitoring (KePIM) (Scek), and the Community Literacy Project Nepal (CLPN) (Maddox and Chitrakar).

<sup>2</sup> In this report, the terms 'measurement' and 'assessment' are not used interchangeably; 'measurement' has been used to refer to a more quantitative, statistical approach, whereas 'assessment' is used in a wider sense to encompass both qualitative and quantitative approaches to assessing literacy. Further discussion on this point can be found in Kell and Rogers' Background Paper.

awareness of the characteristics of other approaches, and the strengths and weaknesses of both. This report is seen as a means of recording the discussions and taking them forward.

### **Constraints:**

The quantitative large-scale measurement approach currently dominates heavily in international agencies and at the national level, despite the shortcomings of past literacy surveys. Efficiency and cost considerations are also crucial constraints on methods of assessment of literacy. How data is presented is as important as the method used. Large amounts of international and national resources and jobs depend on the figures produced by large-scale surveys, and quantitative measurement is often made a requirement. In the past, the media have seized on the apparent comparability of international figures and as a result large-scale surveys can be hugely influential. It is therefore essential that literacy data is reported along with its limitations, rather than as a single figure per country without outlining problems with comparability. However, agencies that carry out literacy assessments are not always able to determine how the data is later presented and used.

### **Key Results and Conclusions of the Seminar:**

Ethnographic and qualitative approaches were seen to be able to contribute to the existing assessment of literacy in three main ways: a) through influencing policy on assessment; b) in large-scale assessment and c) at the micro-scale of assessment. These are each elaborated in the body of the report but are summarised here as follows:

#### **a) Influencing policy on literacy assessment:**

Ethnographic and qualitative approaches were thought by several participants to be able to influence policy discourse directly, such as through independent research and assessment practices separate from large-scale measurement approaches. Ethnography can contribute as a critique, to illuminate the limitations of measurement processes as well as building an ethnography of measurement. Ethnographers therefore need not limit themselves to contributing to policy through incorporation into or supplementation of measurement approaches.

#### **b) Contribution to large-scale assessment:**

Ethnography can contribute to large-scale assessment of literacy/ies in the following ways:

1. By improving the quality and meaning of assessment data: including rigour and cultural sensitivity regarding test constructs and test items through:
  - helping to get the construct 'right' and making it useful - to develop concepts at the level of modelling and variables;
  - integration at the initial design stage to identify and systematise meaningful (and valid, reliable) indicators and descriptors;
  - incorporating concepts of and questions on both school and non-school-based (everyday) literacy practices and contexts (such as 'How do you communicate with relatives or friends who live at a distance from you?');
  - providing an opportunity to integrate literacy and numeracy aspects (as in many everyday literacy activities);
  - improving rating scales for literacy assessment; and
  - deriving items in item response tests from descriptors that are meaningful to learners and teachers.
2. By providing additional background information:
  - identifying the contexts in which literacy learning and practices/uses take place including in which languages (towards best practice); and
  - providing data on attitudes in local contexts such as to reading or language use e.g. sending letters to distant relatives (cf Kell).

3. At the interpretation and analysis stage:
  - by interpreting results in the context of local realities (within schools and learning programmes and outside);
  - by shedding light on how closely results of the survey resemble findings of qualitative data on literacy learning and outcomes (this confers the advantages of a multi-method approach - triangulation); and
  - by assisting the process of turning item scores back into descriptors that are meaningful to learners and teachers.
4. At the presentation stage:
  - by making evident the limitations of statistical measurements for comparison within and across countries
5. By conferring the advantages of a multi-method approach according to the purpose of assessment:
  - identifying factors around process (formative assessment) of literacy learning and participation. The choice of method can depend on whether the purpose of a particular assessment is summative (where quantitative methods are required) or formative (where description can be used).

**c) Ethnography and the micro-scale of assessment:** Ethnography can provide data for the following purposes:

- to inform policy and programmes on what literacy practices and activities people are carrying out in order to increase the relevance of assessment and thus to increase motivation, so that assessment builds on what people already know;
- to facilitate ‘ownership’ of assessment processes and changes;
- to add meaning to and to supplement quantitative monitoring and evaluation processes at the local level; and
- to provide qualitative insight into ‘participation’, ‘empowerment’ and equity aspects in the assessment of literacies.

Given the current dominant influence of cost effectiveness and accountability considerations, distinct types of assessment are likely to be required to satisfy the differences in scale, agencies and actors involved. This does not mean that ethnography must be forced to translate its findings into quantifiable and measurable results. Participants argued that the different approaches or perspectives *cannot* be lined up as a continuum of literacy assessment due to the differences in terms of perspectives, concepts and methods, although they are not mutually exclusive as research orientations (Leung; Kell, from comments). This seminar has shown that ethnographic and qualitative methods can contribute in key ways to the quality of literacy assessment procedures. It has also shown that conventional, statistical methods to assessing literacy do not necessarily prohibit the incorporation of the multiple view of literacy, and that the failure to do so may be a result of resistance to diversity (Freeland, from notes) rather than purely technical considerations.

#### **Future Directions:**

The seminar produced possibilities for moving forward in two main areas:

##### The LAMP programme:

A group formed around the LAMP programme with the following aims in the future: to carry out an ethnographic study of the LAMP process during the piloting stages; to contribute with ethnographic input into the programme itself, helping to define and clarify constructs and items used for assessment, to balance the statistical methods of LAMP, and to refine the methodology used (Freeland; Imperial, from comments).

It was thought that LAMP may be useful in developing countries as an alternative to national surveys for the purposes of larger educational and social planning, if methodological concerns and philosophical underpinnings could be agreed by participating countries (Imperial, from comments). It is likely however that it will provide different figures (usually lower) than the existing modes of assessment, and these may have political implications.

Multi-lingual literacies:

A group that formed and ran throughout the length of the seminar agreed to maintain contact in order to take some of the issues and discussions further with a view to organising a conference on this theme in the future (Freeland, from comments).

## **MAIN SEMINAR REPORT**

### **Background and format of seminar:**

The Uppingham Seminars are a series of conferences held at irregular intervals (normally once a year), organised by Alan Rogers together with a Steering Committee that includes Anna Robinson-Pant and Brian Street; they are normally held at the Falcon Hotel in the small market town of Uppingham in the UK, but other locations (including overseas) have also been used. The Uppingham Seminar for 2003 was held over three days at the end of November 2003.

This year's seminar was attended by 32 people with experience in aspects of literacy and numeracy learning and assessment at the local, national and international levels, including research, development, and participatory monitoring and assessment. Participants encompassed countries from the 'South' and the 'North' including the UK, USA, Germany, many countries in Africa (South Africa, Malawi, Mali, Botswana, Uganda and Kenya), India, Pakistan, and Nepal, Mexico, the Philippines and New Zealand.

The Uppingham Seminars are unlike conventional conferences in that they have an informal structure. With no formal giving of papers, the seminar consisted of plenary sessions and smaller group discussions. The seminar was provisionally timetabled into sessions before we began, but beyond that there was no fixed format; the structure was allowed to unfold as the seminar progressed. A background paper and additional papers by several of the participants were sent out before the event to guide discussions, and during the seminar several people gave short presentations with one person chairing each discussion. It was an opportunity to listen to and share knowledge and experience, and to participate in a wide-ranging discussion of ideas and concepts to do with literacies, in a format and depth that would not normally be possible within a conventional conference.

The papers sent out beforehand on which presentations were based were as follows: A background paper on Measuring Literacy (Cathy Kell and Alan Rogers), the Literacy Assessment and Monitoring Programme (LAMP, UNESCO) (Bénédicte Terryn), Psychometric testing and Item Response Theory (IRT) (Constant Leung), the Equipped for the Future (EFF) approach (Regie Stites), Multi-lingual literacies in Vanuatu (Kara Jackson), Numeracy as a social practice (David Baker), the National Adult Literacy Agency (NALA), Ireland (Juliet Merrifield in her absence due to illness), Kenya Participatory Impact Monitoring (KePIM: Aues Scek), and the Community Literacy Project Nepal (CLPN: Bryan Maddox and Roshan Chitrakar).

### **The Topic: Measuring literacy**

The aim of the seminar was to try to find common ground between ethnographic methods and more conventional measurement procedures for assessing literacy; and to seek out and document ways that ethnography can contribute to measuring literacy. Integral to the discussion of how to assess literacy is the nature of literacy itself.

The starting point was the concept of multiple literacies and the consequences of this view for measurement of literacy: if literacy is a social practice with multiple dimensions, i.e. literacies, how does this influence assessment?

Substantial criticism has been levelled at existing modes of literacy measurement and assessment and associated data by ethnographers. Ethnographers have, in turn, been challenged to make a positive contribution to ways of measuring literacy. The seminar and this report are therefore an attempt to seek out and document ways that ethnographic methods can contribute to measuring literacy.

## **Rationale behind the choice of topic:**

### Criticism of Methods of Measurement:

The topic resulted from concerns by a range of people in the field regarding the adequacy of current modes of measurement and assessment of literacy. This includes international agencies such as UNESCO and DFID on one hand, and qualitative researchers and ethnographers on the other.

The Millennium Development Goals (MDGs), Education For All (EFA) initiatives and forums such as Jomtien and Dakar have resulted in pressure to produce measurable statistics that can be compared across countries and over time. This applies across a range of disciplines (Bernstein et al 1992); literacy is no exception. Achieving a 50% improvement in levels of adult illiteracy by 2015 is the fourth EFA goal. Despite doubts about a direct link between literacy and development (Rogers 2001), the youth literacy rate and the ratio of literate females to males are also used as indicators for progress on several MDGs such as formal schooling, health, HIV/AIDS, poverty, gender equality and other basic human needs (Terry, seminar paper).

Agencies such as UNESCO and DFID have themselves expressed concerns over the validity, reliability and comparability of data and statistics gathered so far on literacy. To date, data for large-scale statistics has been based on indirect measures such as individuals' self-declaration, or 'proxy' indicators such as number of years of schooling (UNESCO Institute for Statistics (UIS) leaflet), or on standardised tests similar to those found in primary school education (Kell and Rogers, background paper). The International Adult Literacy Survey (IALS) showed that indirect measures such as number of years of schooling are unreliable and inadequate; some respondents had a literacy level above that of their official educational level and some below (UIS leaflet).

The comparability of IALS data at national level has been questioned from within the statistical disciplines themselves owing to flaws in the standardised sampling and procedure:

“... all comparative analyses across countries should be interpreted with due caution. In particular we recommend against the publication of comparisons of overall national literacy levels. We consider any rankings of countries based on such comparisons to be of dubious value given the methodological weaknesses...” (Kalton, Lyberg and Rempp 1998:4, cited in Hamilton and Barton 2000).

Qualitative and ethnographic criticisms of literacy measurement maintain that literacy is a complex phenomenon which cannot easily be described by a dichotomy of 'literate'/'illiterate' on an individual level, by a single statistic at the country level, and by measuring literacy in the dominant language only according to an internationally standardised view of what constitutes 'literacy' (Hamilton and Barton 2000; Street 1996). Some of the criticisms come from proponents of the New Literacies Studies (NLS) view, which sees literacy and numeracy as social practices with a multiple nature, differing significantly according to the context (Street 1993; 2001; Gee 1990). This includes, for example, multi-lingual literacies, non-prose literacies and religious literacies. Definitions (or constructs) of literacy are always power-laden, as they determine what is included and what is excluded. According to this view, the *dominant* concept of literacy in the dominant language does not reflect an intrinsic superiority but *results from specific historical social and cultural processes which have made this view hegemonic* (Freeland, from notes). The NLS view holds that large-scale comparisons at country level of literacy using one figure are inappropriate since they are not contrasting like with like, even without the flaws in the standardisation procedures mentioned above (Hamilton and Barton 2000; Baker, seminar paper).

## Multiple literacies?

Debates on measurement of literacy exist in the context of ambivalence in international discourse on the nature of literacies, that is, whether literacy is *multiple* or *single*:

“Literacy is no longer seen as a singular concept, but rather as plural ‘literacies’. These literacies differ according to purpose, context, use, script, language and institutional framework... Within any community there will be a range of *literacies* – understanding what these are and how they are structured is important for negotiating whether and how *literacy* might be acquired.” (UNESCO Global Monitoring Report 2002: 60, emphasis added).

Despite the presence of a corpus of work on *multiple literacies*, approaches used to measure ‘literacy’ for statistical comparison at the national level have so far not taken this into account. Instead, it is assumed that literacy skills can be detached from social practice and measured without considering questions such as what is being read and/or written and for what purpose, in which language, together with the level of fluency and confidence (Kell and Rogers, background paper).

A number of efforts to reform literacy assessment are being made. These include:

- contextualising instruments of testing in terms of language, and using, for example, culturally appropriate texts (e.g. gathering of background information in LAMP);
- increasing and refining the number of ‘levels’ of ‘literacy’ (e.g. as in the use of household surveys);
- separating different ‘fields’ of literacy such as document, prose, numeracy, and different competencies, usually involving the (contested) idea of an emergent hierarchy according to perceived equivalence with formal schooling;
- widening the ‘scope’ of ‘literacy’ to include basic learning competencies or life skills which do not require literacy in a dominant language;
- calling for the supplementing of quantitative data with qualitative data.

(adapted from Kell and Rogers, background Paper)

## Approaches: measurement vs ethnographic approach

The participants of the seminar came from a range of standpoints; to an extent differences in approach reflected disciplinary boundaries (such as psychology, psychometrics, linguistics and quantitative/mathematical disciplines contrasted with anthropology, sociology, and socio-linguistics) and the different paradigms within which academics and practitioners work, particularly constructivism contrasted with positivism which runs across both quantitative and qualitative approaches.

Views on how to measure literacy were inextricably linked with views on literacy itself. Broadly speaking, two main strands emerged with two sets of discourses and terminologies associated with them. The first was the measurement and large-scale assessment approach used to assess literacy at country level by agencies such as UIS and the World Bank. This approach primarily involves the use of quantitative methods drawing on item response theory, psychometric testing and statistical techniques to measure *individual* competencies and outcomes of learning, and to produce large-scale statistics at a given time such as is used in a census. This methodology is broadly associated with a conceptualisation of literacy as taught in the basic skills, linguistics and technical approaches and emphasises prose or essay text literacy<sup>3</sup> (UIS leaflet; Street 1994). The underlying assumption is that skills can be scored in a linear progression, and normally involve assessment in the dominant language of the country.

---

<sup>3</sup> alternatively described as ‘discrete items of language knowledge at the sentence level for transactional purposes’ (Constant Leung, from comments).

The second main standpoint is an ethnographic approach. This involves using primarily ethnographic and qualitative methods to assess outcomes of learning and to provide a more in-depth picture of the literacies processes, practices and uses of people in everyday life (Street 2001). The ethnographic approach also seeks to examine less dominant or vernacular literacy practices (Barton and Hamilton 1998) which ordinarily do not show up on literacy statistics and are therefore often not counted as literacy. This includes, for example, multi-lingual literacies, religious literacies and economic literacies (Baynham 1993; Herbert and Robinson 2001).

Many of the participants concerned with ethnographic methods at the seminar also represent the New Literacies Studies group (NLS), a school of thought that has come to place literacy learning, practices and uses as embedded in social practice and as therefore 'ideological' (Street 1993). This approach has grown out of interdisciplinary studies of literacy, drawing on socio-linguistics, anthropology and sociology. According to this school of thought, literacy practices are often carried out and mediated in groups, subject to social inequalities, as a result of a socialisation process in particular social and cultural contexts, rather than as purely 'autonomous' individual skills (Street 1993). The New Literacies Studies (NLS) considers that literacy and numeracy skills are often taught and measured by decontextualising them, thereby taking away much of the situational clues that people rely on to give meaning to texts and literacy practices.

### **The Problem of Scale**

It became clear that at the seminar there were differing rationales behind assessing literacy between the different interested groups and stakeholders. These represented differences in scale and power: at the macro-scale, from powerful international agencies funding literacy programmes and concerned with accountability; at an intermediate scale, from implementers at programme level; and at the micro-scale, from the teachers and learners themselves. A vital question emerged: can these differing interests be satisfied by one method of assessment? Or will different measurement processes be called for?

### **Measuring Literacy: why and for whom?**

Over the course of the seminar, key questions relating to the scale and power dimension were addressed: Why are we measuring literacy, and for whose benefit? Is the key aim to satisfy accountability concerns of agencies, and to provide aggregate results for the purpose of comparability? Or is the main reason to provide feedback to learners and to aid participation in the learning and teaching process? Both approaches stated that the aim of assessment is to inform policy and improve implementation, learning and teaching, and to improve equity. However, the means by which this is best done is contested.

The primarily statistic-based International Adult Literacy Survey (IALS) had a striking influence on literacy provision in a range of countries (Crowther, Hamilton and Tett 2001). However, measurement for the purpose of accountability was not necessarily seen by participants at the seminar to be congruent with equity and the needs of the learners. Vincent Greaney of the World Bank highlighted that measuring for accountability can improve assessment procedures significantly, although it can also have a detrimental effect on learner motivation.

In Ireland, the National Adult Literacy Agency (NALA) has developed a performance-based assessment framework which is not intended for reporting results in an aggregated way or as a mechanism for accountability (Merrifield, seminar paper). It was argued that ethnography and the New Literacy Studies approach do not necessarily need to bow to the pressures of measurement. Nevertheless, the literature on NALA in Ireland states that in practice, despite widespread resistance from within, some form of measurement will have to be developed or else imposed using other methods in the future for reporting purposes.

### **Individual and social aims: exposing a false dichotomy?**

The distinction between the individual and the social dimension was crucial during the discussion about the purpose of assessing literacy. Should we be assessing specific literacy skills as useful in themselves to individuals, or assessing them with a view to achieving various other aims including employment? If literacy programmes are implemented with the aim of promoting the social dimensions of ‘development’, ‘empowerment’, ‘participation’, ‘poverty reduction’ and ‘social justice’ within populations and communities, should we be measuring indicators of ‘participation’ and ‘development’ rather than literacy? Similarly, if ‘functional’ literacy is the model used, as it is in many developing countries, should we be assessing functional outcomes such as health benefits and increased agricultural productivity? (Kisira; Imperial, from comments during seminar).<sup>4</sup>

Therefore, by measuring literacy skills per se, are we concentrating on the *means* to the exclusion of the *ends*? Furthermore, should we be ensuring that the measurement or assessment process itself enhances participation (Chilisa, from notes on discussions)? These differences in approach and aims were held in tension through much of the seminar discussions.

One of the participants stated that the standardised basic skills approach to literacy is privileged over the NLS view by the learners themselves (Stites, from notes on discussion). Nevertheless, it has also been documented that enrolment in literacy programmes that use standardised literacy approaches is not as high as this statement would suggest (Kell 2003), while non-completion rates are very high (Rogers, contribution in seminar). However, the privileging of the ‘schooled’ literacy or basic skills model over the social view raises a more general political question: is this merely the dominant ideology favouring the view which allows easy measurement and which does not ask awkward questions about social inequalities (see Bernstein et al 1992)?

The seminar discussions largely tended towards a dichotomy between basic skills and the measurement of literacy skills per se on the one hand, and qualitative approaches to assessment and the social view of literacies on the other hand. Both originate in developed countries. It was felt by one participant that this was imposing a view of literacy from outside and not fully taking into account the voice of developing countries (Robinson-Pant, from notes on sessions). (Recent work by Brandt and Clinton (2002) has also stressed that an alternative approach can be found between the two positions in the dichotomy.<sup>5</sup>)

---

<sup>4</sup> Many literacy programmes in ‘developing’ countries are integrated to some extent into wider skills programmes or adult basic education and training; the Philippines is one example: “To us, functionality actually approximates the essence of ‘multiple literacies’ and even oral competencies which may be a big impact of one’s being able to learn how to learn despite inadequacy of schooling and one’s exposure to mass media. This is why mass media is a component of our FLEMMS [Functional Literacy and Exposure to Mass Media Survey] programme [in the Philippines]. Perhaps developing countries are just being pragmatic when functional literacy is integrated with skills programmes. *Literacy should not be a stand-alone goal and process but should be pursued because of the benefits needed for one’s day-to-day existence.*” (Imperial, from comments, emphasis added).

<sup>5</sup> Brandt and Clinton (2002) have indicated that middle ground can be found between the technical, individual view of literacy and the social, multiple approach to literacies. They argue that the NLS approach grew out of a response to the technical, individual view, and in so doing made valid criticisms of the technical view, especially with regard to cultures which are based more around oral culture than literate culture. However, they claim that the NLS approach has moved too far the other way and has lost sight of the material aspects of the process of learning literacy. Both Baynham, (with a number of authors in a special issue of *Language and Education*, 2004) and Street (2003) have responded to these arguments pointing out that ethnographies of literacy do always already take account of the links between the local and the global: ethnography, by definition, is about comparison,

### **The question of validity**

One of the major elements of contestation of methods of assessment in literacy statistics was over what is considered to be valid assessment. Conceptualisation of validity differed between the various points of view and appeared to be incompatible in some respects (Teasdale and Leung 2000). Validity in quantitative approaches is concerned with the extent to which an indicator - in this case a test item - accurately measures the concept it is designed to measure. In contrast, according to Teasdale and Leung, validity for ethnographers and qualitative scientists is concerned with whether the theory or model adequately reflects practices in 'real-life' contexts or not. However, according to Stites and the EFF approach, the validity of performance (alternative) assessment can be achieved by clearly specifying the content domain (or construct) to be measured, and structuring tasks to ensure the production of sufficient evidence to support assessment interpretations (Stites, seminar paper). Validity is further discussed below (see also Appendix).

It was pointed out that both processes involve modelling and the development of constructs, and that this needed further consideration. Kell argues that, while the ethnographic approach puts the modelling of its constructs at the centre of its enterprise, the 'naturalisation' of the constructs in the form of statistics can conceal the culture-specific and subjective processes guiding the origins of the constructs from which the figures are derived in the first place (Kell, from comments).

Critical arguments from ethnographers and the New Literacy Studies (NLS) group have questioned the validity of commonly used quantitative methods such as Item Response Theory and psychometric approaches to measuring literacy, on the grounds that these approaches miss out crucial situational and contextual information and treat differences in culture as a source of bias (Hamilton and Barton 2000). According to this view, the test items in IALS, rather than representing 'real life' practices, in fact assessed familiarity of respondents with a kind of transnational culture of 'test literacy', biased recognisably towards US-based culture (ibid:385). Instead, ethnographers in the NLS advocate the use of qualitative approaches to highlight the diversity of actual literacy practices and uses by people who would normally be described by conventional literacy statistics as 'illiterate' (cf Street 1996).

Ethnographers' conceptualisation of validity has in turn been criticised by quantitative social scientists as being inadequately developed or empirically justified; as framing validity in terms of a procedure (in performance-based assessment) without considering the problematic nature and sources of bias in assessment where teacher judgment is a deciding factor. Validity in qualitative approaches has also been criticised for borrowing inappropriately from concepts of validity in psychometric approaches for conditions where some of the basic assumptions do not hold, while distancing themselves from such approaches (Teasdale and Leung 2000). An in-depth discussion of validity therefore shows that the two approaches cannot simply be combined without careful consideration.

### **The notion of 'levels' of ability in literacy**

Measurement and statistical approaches place learners on a scale of ability, assumed to be linear. While this is a step forward from a 'literate'/'illiterate' dichotomy, empirical evidence has shown (with two separate instances detailed later in the report, one by UIS and the other by CLPN) that such scales are not necessarily linear. Learners can master a competence thought to be at a 'higher' 'level' without ostensibly having mastered the corresponding 'lower' competence or 'level' of ability (UIS leaflet; Maddox, from presentation).

---

not simply isolating the 'local' (Street, from comments). This seminar created a space for dialogue between the various views to take place.

The tests used in the large-scale literacy survey IALS were “designed to ensure a broad spread of responses across an arbitrarily fixed set of five levels”, allocating a significant proportion of people to each of the five levels (Hamilton and Barton 2000:384). This means that the ‘levels’, being arbitrarily determined, bear no relation to the actual tasks that people perform in their everyday lives. In addition, research has indicated that the difficulty of reading any text is not inherent in the text itself (there is no such thing as an ‘easy reader’) but lies in the experience of the reader, so that uniform tests can never measure the individual abilities of each participant.

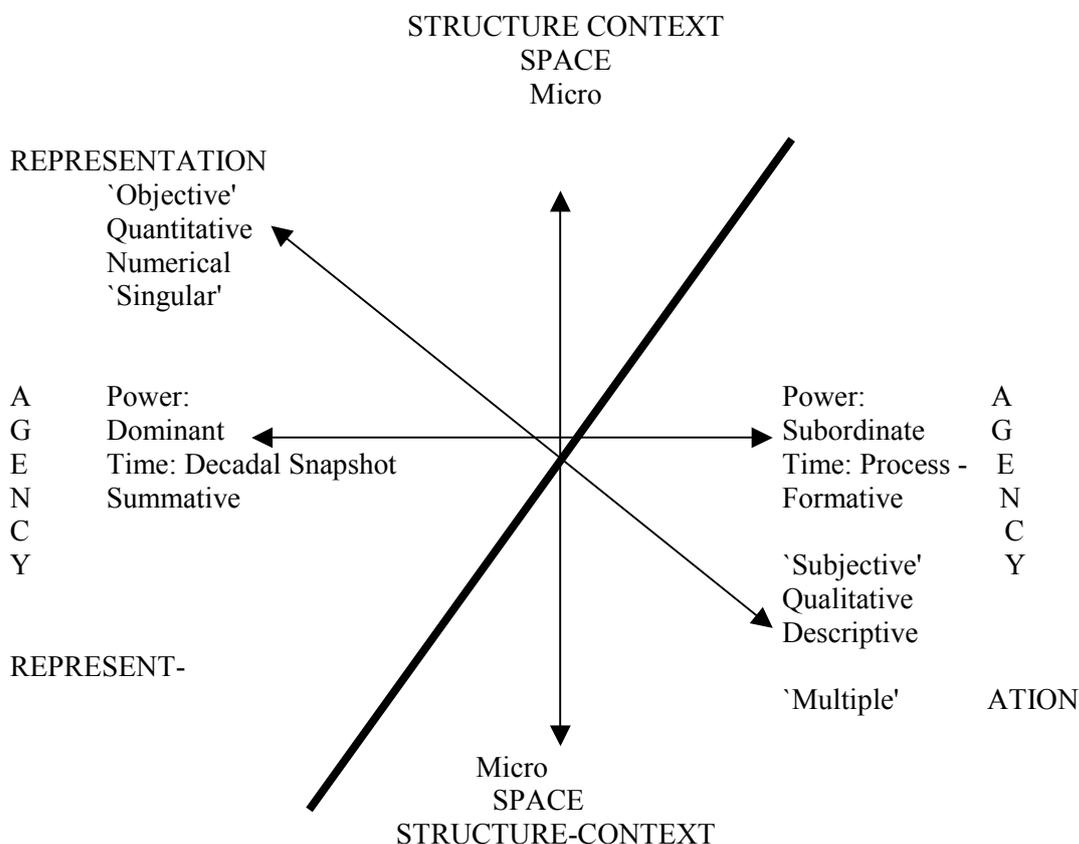
### Making comparisons in literacy

Tension between participants was apparent over how and whether literacy assessment results can be compared; this is the main aim of producing large-scale literacy statistics. The tensions reflect the high stakes that are involved: comparative statistics have the power to shape resources on an international scale. Some participants at the seminar however contested the validity of making any large-scale comparisons. However, this has to be examined in light of the fact that ethnographic work is always modelled and therefore attempts to make some kind of comparison. It became clear that many of the differences between methods were at the level of modelling and the particular models used to drive literacy assessment.

### Modelling

A number of models were put forward during the discussions, some arising out of small group discussion and some in the plenary sessions.

1. The following diagram was produced by Denzil Saldanha in one of the small group sessions, illustrating the relationship between different types of assessment and the views of literacy/-ies:



## 2. Ideal type models:

Brian Street asserted that models in social science are often based on the ideal type model such as Weber's model of bureaucracy (cf Giddens 1997:287 see also Appendix). Models propose a heuristic that tries to capture some of the most salient features of a system or process, simplified ways of capturing the 'messiness' of social reality, for example, the model of capitalism. The aim of such models is to allow prediction with the smallest number of indicators rather than to be right in every detail. They are not intended to be able to describe all aspects of society. As Weber argues:

'An ideal type is formed by the one-sided *accentuation* of one or more points of view and by the synthesis of a great many diffuse, discrete, more or less present and occasionally absent *concrete individual* phenomena, which are arranged according to those one-sidedly emphasized viewpoints into a unified *analytical* construct (*Gedankenbild*). In its conceptual purity, this mental construct (*Gedankenbild*) cannot be found empirically anywhere in reality' (Weber 1949:89).

For Weber and many social scientists, an ideal type construct is a heuristic for bringing together key features of a social system or process. It is not, then, itself an empirical description although working with 'models' in this sense enables researchers to describe and analyse aspects of empirical reality. The concept of literacies as multiple is such a model that has facilitated numerous ethnographic accounts of literacy practices. Such models are always representations of reality, which is always rather messier and complicated than a model of, say, 'academic literacy practices' or 'religious literacy practices' can be. The situation is further complicated in literacy assessment because this involves modelling what is *already* a *representation*: so the constructs used for testing and the extrapolations from them are already secondary representations, a feature that gets forgotten when they are used to make generalisations and comparisons across countries.

But model building is not simply the preserve of social science, to be marginalised as 'soft' compared with the realities of the 'hard' sciences, for scientists and mathematicians also work through modelling (see quotes from Gowers in Appendix). As Gowers states, in language similar to that used by Weber: 'Mathematicians do not apply scientific theories directly to the world but rather to *models*. A model in this sense can be thought of as an imaginary, simplified version of the part of the world being studied, one in which exact calculations are possible'.

These accounts of model building suggest that the distinction between qualitative and quantitative researchers could thereby be seen as less divisive than the distinction between *positivist* and *interpretive* (constructivist) paradigms or epistemologies. Both qualitative and quantitative researchers, as Weber and Gowers suggest, can be seen as constructing models that 'accentuate' a 'version of the part of the world being studied'. In the seminar, the common recognition that both social literacy theorists and assessment experts were working with such models helped to bring the two parties together and to provide a bridge for the conversations.

## 3. Models 'of' and models 'for':

Cathy Kell, Constant Leung and Jane Freeland (with reference to Goffman 1974) highlighted the differences between models 'of' and models 'for' used in literacy assessment. Models 'of' are constructed to describe what is happening, in this case in an assessment of literacy/ies. Like ideal type models, models 'of' are a simplification of the situation. Models 'for' are not necessarily based on models 'of' and often incorporate value aspects, that is, a vision of what we would like the situation to be. As Weber puts it: we should emphasise that

the idea of an ethical *imperative*, as a ‘model’ of what ‘ought’ to exist, is to be carefully distinguished from the analytical construct, which is ‘ideal’ in the strictly logical sense of the term. In Kell’s terms, the ethical imperative is a ‘model for’, whilst Weber’s logical construct is a ‘model of’.

Agencies and governments favour models ‘for’, where central political control or intervention is seen to be necessary in assessment. Such models are not necessarily based on empirical evidence but are driven by the agency’s vision and theory of how to achieve its goals. In educational systems where democratic decision-making and diversity are a greater feature, assessment of literacy will tend to be based on models ‘of’ and theories which are non-hegemonic (Freeland, from comments). But economic, political and ideological interests play a decisive part in the formation of models for literacy assessment (Leung, from notes). A simple opposition between ethnography as using models ‘of’ and statistical constructs as using models ‘for’ would also be misleading. Ethnography and statistical constructs are both theory driven and use models ‘of’. The ‘bottom up’ (and model ‘of’) approach has wide implications since the results of such an assessment process may not be what policy makers expect or want.

#### 4. Diversity and multi-lingual literacies:

As in Denzil Saldanha’s diagram above, qualitative and ethnographic methods were favoured (in the small group discussion) for assessing multi-lingual literacies, owing to the complexity of the literacy practices and literate environments. Nevertheless, statistical techniques have evolved to deal with multiple variables, so the barrier is not in fact technical but one of resistance to diversity by institutions (Freeland, from notes).

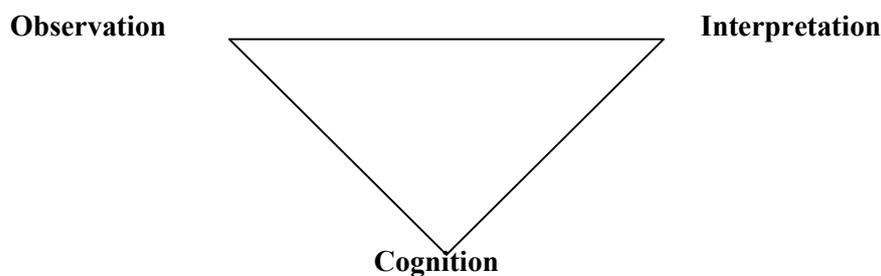
#### 5. Qualitative and quantitative assessment:

Regie Stites pointed out that the USA’s National Research Council (NRC) identifies three basic elements of good assessment:

- **cognition** – the definition of the aspects of achievement to be assessed
- **observation** – the definition of the tasks used to collect evidence about performance ability
- **interpretation** – the methods used to analyse the evidence resulting from the performance on the tasks

These three elements are portrayed as an “assessment triangle” (see below) to emphasise the interrelationships among all three elements. According to the 2001 NRC report, no assessment can be designed and implemented without consideration of each of the three elements in the triangle. Making each element explicit and making sure that all three elements work well together is the key to good assessment (EFF, forthcoming).

#### The Assessment Triangle



The assessment triangle is a heuristic that is useful for making explicit the framing assumptions of any approach to educational assessment (whether qualitative or quantitative). The chief value of this heuristic is that it forces us to make our assumptions explicit about what we are measuring (the cognition/ learning/ performance model); how we structure tasks or situations (test items/problems) to allow us to observe the knowledge or skills in question (the observation model); and how we interpret the resulting response/performance/work product. Every test is built out of these three elements (Stites, from comments).

Stites highlighted the process and order in which the three parts of the assessment triangle normally take place: the primary consideration in IRT is the interpretation and analysis stage and making sure this stage is manageable. The other two aspects of the triangle (cognition and observation) are developed out of the analysis considerations. This is therefore a type of model 'for'. It was felt by Stites, however, that the assessment triangle as a model could incorporate both empirical information into the construct and more reflection on the theoretical assumptions underpinning the construct, thus making it more of a model 'of' (from notes on small group discussion).

Item Response Theory (IRT) is a scaling method – and fills the interpretation corner of the assessment triangle. However, IRT places constraints on the ways we are allowed to conceptualise literacy skills and the ways we can design tasks to measure literacy skills. In particular, IRT can only be used as a scaling method for a *unidimensional* measure of literacy. That means that IRT is basically *incompatible* with *multidimensional* understandings of literacy (such as in NLS), which is a major weakness of the theory (Stites, from comments).

#### 6. 'Levels' and profiling:

Bénédicte Terryn outlined that LAMP aims to choose a model according to the particular characteristics of the data (therefore making it to some extent a model 'of'). The particular configuration of the data is to determine whether an approach with 'levels' of literacy or a more complex '*profile*' approach will be used to analyse the data. Nevertheless, the LAMP programme uses a framework of a 'filter test' and two-level approach (elements of a model 'for'), which distinguishes between a 'low level' of literacy and a 'higher level' of literacy (Terryn, seminar paper).

#### 7. Epistemology:

Participants were divided over whether the differences extended to the level of epistemology (fundamental theories of knowledge) or not (Street; Robinson, from notes). Nevertheless, psychometric approaches are held to be incompatible with values underlying particular pedagogies and epistemologies such as constructivism<sup>6</sup>, and with socialisation and 'scaffolding' approaches or integrated curricula that are found in alternative approaches (Teasdale and Leung 2000).

### **Literacy and Orality:**

During discussions, Judy Kalman outlined recent developments in research on cognitive learning which highlight that reading and writing, rather than being the building blocks that we learn at the beginning, are actually learned after a prolonged period of socialisation, discussion and contextualisation, using orality (from notes on discussions). For instance, children usually come into contact with words first through speech, later through having stories read to them. Only after that do they start to read and write words for themselves<sup>7</sup>.

---

<sup>6</sup> Constructivism holds that objects in an area of inquiry (social science) are not there to be discovered but are invented or constructed (Mautner, T. (ed.) 1996, reprinted 2000 *The Penguin Dictionary of Philosophy* London, Penguin Books:111)

<sup>7</sup> A musical analogy serves to illustrate the literacy/orality issue. Musicians can and do become respected performers and teachers without reading music in the dominant (classical) form of notation,

Focusing on the decontextualised skills of reading and writing, therefore, is trying to leapfrog this process and start at the end, not at the beginning. Speaking and listening are often implicit or tacit in the school curriculum whereas written skills are more bounded and focused - such as in the 'literacy hour' in the national Literacy Strategy in England, or 'reading' classes (Leung, from notes). This affects how they are assessed.

The particular emphasis on literacy as opposed to orality in communication differs significantly across cultures in different countries and even within the same country (Brandt and Clinton 2002). Such differences and the correlations between predominantly oral cultures and perceived under-development have in the past led to causal relationships being asserted (ibid.) and unrealistic claims being made for literacy (Rogers 2001).

Baynham (1993) asserted that the dichotomy between literacy and orality breaks down under close scrutiny. Socio-linguistic analysis has shown that reading and writing in context closely resemble the functions of speech and are therefore just one aspect on a continuum of communication to address various human needs (Brandt and Clinton 2002). Baynham's (1993) framework of 'mode-switching' and 'code-switching', together with the different roles played in literacy practices (Kell 2003) is useful to illustrate the sometimes arbitrary categorisation of 'literacy' and 'illiteracy' which is effected by privileging particular literacy aspects over orality. Code-switching involves crossing between different languages; it commonly occurs in literacy practices in multi-lingual environments, but it also includes crossing between registers such as formal or official and informal or colloquial language. Mode-switching involves changing between different types of language use such as oral and written.

Kell (2003) reports on a range of work undertaken by Freebody, Luke, Green and Lankshear amongst others which distinguishes between roles carried out by different participants in literacy events and practices such as 'code-breaker', 'text user' and 'text participant' and 'text analyst'. Research has shown that people who are not proficient in code-breaking (deciphering print) are nevertheless able to fulfil other roles of literacy mediator, text user and text participant, where texts are used and discussed in particular social contexts. The roles of text user and text participant are often assumed in conventional discourse to come after the code-breaking stage (Kell 2003). The roles of literacy mediator, text user and text participant involve proficiency in 'secondary discourses' (Gee 1990: 151-2) which typically rely on a combination of oral elements as well as written text. Such literacy uses and roles are examples of literacy practices carried out by people officially categorised as 'illiterate' by the dominant culture and who may have internalised the dominant view to classify themselves as such. Privileging reading and writing over oral forms of communication has equity and social justice implications (from presentation, David Baker).

### **Multi-lingual literacies:**

The theme of multi-lingual literacies was a key concern in plenary sessions and was developed throughout the whole seminar by one group of participants. Since I did not attend the meetings of this small group, I have drawn heavily on participants' notes and comments on what was discussed. This group was concerned with the particular aspects and complexities of literacies in multi-lingual environments and the consequences for measuring literacy.

---

both in classical and other forms of music. Many would argue that the aural and performance elements come first, the 'reading' and notation afterwards. In addition, alternative forms of coding exist, such as 'ABC', where music is notated using letters. It can be argued that to assess musical competence privileging the ability to understand and write classical musical notation rather diminishes the range of competencies involved.

The use of particular languages and literacies in specific contexts is based on historical and cultural factors (Herbert and Robinson 2001), and the issue of multi-lingual literacies links with considerations of power, empowerment and comparability. The relationship between languages and literacies in contexts where people use more than one language in their daily lives is complex. It cannot be mapped in terms of one type of literacy in one language. Instead, it is necessary to observe patterns of language use and how they are distributed across different communicative events, including oral and written forms (Freeland, from her notes on multi-lingual literacies group). It is also necessary to examine for each particular context what determines a choice of language for what purpose in a particular social setting, what power differentials are involved, and how literacies relate to these patterns.

Analysis of such configurations shows that literacies in different languages cannot simply be compared, since they do not have identical status, that is, they do not 'empower' groups in the same way or to the same extent with regards to the dominant culture. Therefore, combining them to make one single figure is misleading. For example, Shirin Zubair pointed out that religious literacies such as Koranic literacy in Pakistan is less 'empowering' for women than the dominant literacy, yet it is the literacy which they are most readily allowed access to by males (Freeland, notes on multi-lingual literacies). In multi-lingual contexts, different literacies are linked with negotiating complex and changing identities. One of the participants stated how in India, women who have become literate may be prevented from using their literacy skills through social relations and prescribed or reserved gender roles (Ghose, from notes).

Attempts to integrate multi-lingual literacies into measurement of literacy so far were criticised for still assuming comparability or transferability. The difference was judged to be conceptualised (even in recent surveys such as LAMP) as a matter of technical decoding, lacking any social and political context. This ignores the fact that mastering literacy in some languages may not confer the classification of being 'literate' (Freeland, from notes on multi-lingual literacies).

### **Literacy, development and social justice**

The link between literacy and development is complex and problematic, especially given the unrealistic claims for it in the past (Rogers 2001). Despite this, literacy rates are being used as indicators for much wider development aims as part of the Millennium Development Goals and Education for All. Participants gave examples, from India (Ghose, from notes), and Mali (Puchner 2003) where the benefits of literacy intervention are highly constrained by existing social and gender relations. For example, women who had mastered literacy skills were nevertheless not allowed to use them, and roles involving literacy were dominated by men. Therefore, the question was raised: are isolated literacy programmes (and by extension using literacy as an indicator) the most effective way to tackle social justice and equity concerns? (Puchner, from notes)

### **Efficiency and cost**

In practice, the choice of methods and tools for assessing literacy is highly constrained by considerations of cost and efficiency. Such considerations seem to lead to different choices of method according to differences of scale. At the country scale, statistical quantitative methods are more cost-effective and efficient, especially those relying on proxy measures: ethnographic methods are time-consuming and costly. At the micro-scale, however, quantitative methods such as psychometric testing can be expensive; performance-based assessment is more cost-efficient as the bulk of the assessment is carried out by the teachers themselves (Teasdale and Leung 2000), but this is also open to criticism.

### **Presentation and use of results in literacy assessment**

How data is presented was seen to be as important as the method used. It was thought to be essential that literacy data be reported along with its limitations rather than (for example) as a

single figure per country without outlining problems with comparability (Hamilton and Barton 2000). However, it was pointed out that agencies which carry out literacy assessments are not always able to determine how the data is later presented and used by, for example, the media (Terry, from notes). David Baker showed how modelling processes are cyclical and the original process of developing the constructs should be (but is not always) returned to in the final presentation, so that the socio-cultural and subjective variables used in the modelling can be understood (Kell, from comments).

Several participants highlighted that quantitative statistical methods seem to command an unquestioned authority and hegemonic character that qualitative data cannot challenge; and that statistics create a sometimes illusory impression of standardisation and scientific rigour. The findings of the IALS were influential internationally (Crowther, Hamilton and Tett 2001) and the media seized on the apparent comparability of international figures (Hamilton and Barton 2000; see also the special edition of *International Review of Education* 46 (5) 2000). Large amounts of international and national resources and jobs depend on the simplification of complex situations to simple figures, and governments and other bodies often make such measurement a requirement. Statistics can be interpreted and used in different ways for different purposes and audiences; to emphasise the gravity of the situation (to agencies to try to get more money) for example, or how much a programme has achieved (to show what good results have been produced) (Kisira, from notes). Participants called for greater challenging of such methodologies. Although such questioning was reported to happen within quantitative assessment disciplines (Leung, seminar paper), the contested nature of such procedures rarely seems to filter through when the results are presented to audiences not experts in the field (Hamilton and Barton 2000).

#### **Constraints on the use of ethnography in large-scale assessment**

A number of constraints to the use of ethnography in large-scale assessments were expressed during the seminar. These included:

- the need for summary measures and indicators that can be used across countries to satisfy agency and government requirements (Terry; Leung, from comments);
- the characteristics of the policy context: ethnography's perceived lack of compatibility with models preferred by agencies for literacy assessment, and support or resistance to ethnographic methods, change and diversity within agencies (Greaney, from comments);
- the resource, time and practical considerations and cost effectiveness: this has a potentially overriding influence on actual methods used and prohibits a full investigation into all the 'minority' (non-hegemonic) literacy practices in every country (Terry, from comments). Likewise, ultimately the practice of producing a single score for comparison is likely to stay (Leung, from comments);
- the specific requirements of agencies for objectivity and justifiability of results of assessment, i.e. reporting in terms of quantifiable and measurable characteristics (Greaney, from comments);
- interpretation of the results in the light of local realities is more suitable to a population approach to assessing literacy rather than a sample approach (Greaney, from comments).

#### **RESULTS AND CONCLUSIONS:**

##### **Contributions of ethnographic approaches to measurement:**

The main ways that ethnography was seen to be able to contribute to the assessment of literacy outcomes are grouped as follows: a) contributions to policy discourse; b) in large-scale measurement and assessment (e.g. LAMP); and c) at the micro-level of assessment:

**a) Contribution of ethnography to policy making:**

Ethnographic and qualitative approaches were thought by several participants to be able to influence policy discourse directly, such as through independent research and assessment practices, separate from large-scale measurement approaches. Ethnography can be used to illuminate the limitations and successes of the assessment process itself (an ethnography of assessment) (Kell, from comments). Ethnographers therefore need not limit themselves to contributing to policy through incorporation into or supplementation of measurement approaches.

**b) Contribution of ethnography to large-scale assessment:**

Ethnography was thought to be able to contribute to large-scale assessment of literacy/-ies in the following ways:

1. by improving the quality and meaning of assessment data: including rigour and cultural sensitivity regarding test constructs and test items through:
  - helping to get the construct ‘right’ and making it useful - to develop concepts at the level of modelling and variables (Street, from comments);
  - helping at the initial design stage to identify and systematise meaningful and valid, reliable indicators and descriptors;
  - incorporating concepts and questions on both school and non-school-based (everyday) literacy practices and contexts such as ‘How do you communicate with relatives or friends who live at a distance from you?’ (Leung, from notes);
  - providing an opportunity to integrate literacy and numeracy aspects (as in many everyday literacy activities);
  - helping to improve rating scales for literacy assessment; and
  - deriving items in item response tests from descriptors that are meaningful to learners and teachers.
2. by providing additional background information:
  - to identify the contexts in which literacy learning and practices/uses take place including in which languages (towards best practice); and
  - to collect data on attitudes in local contexts such as to reading or language use e.g. sending letters to distant relatives (cf Kell).
3. by assisting at the interpretation and analysis stage:
  - to interpret results in the context of local realities (within schools and learning programmes and outside);
  - to shed light on how closely results of the survey resemble findings of qualitative data on literacy learning and outcomes (this confers the advantages of a multi-method approach through triangulation); and
  - to aid in the process of turning item scores back into descriptors that are meaningful to learners and teachers.
4. by working at the presentation stage:
  - to make evident the limitations of statistical measurements for comparison within and across countries.
5. by conferring the advantages of a multi-method approach according to the purpose of assessment:
  - identifying factors around the process of literacy learning and participation; the choice of method can depend on whether the purpose of a particular assessment is summative (where quantitative methods are required) or formative (where description can be used).

**c) Ethnography and the micro-scale of assessment:**

Ethnography can provide data:

1. to inform policy and programmes on what literacy practices and activities people are carrying out in order to increase the relevance of assessment and thus to increase motivation, so that assessment builds on what people already know (Street, from comments);
2. to facilitate ‘ownership’ of assessment processes and changes (Street, from comments);
3. to add meaning to and to supplement quantitative monitoring and evaluation processes at the local level (as was the case in the KePIM programme – see Appendix); and
4. to provide qualitative insight into ‘participation’, ‘empowerment’ and equity aspects in the assessment of literacies.

### **Meeting in Collision?**

The aim of the seminar was not necessarily to reach a consensus, but to provide a space for discussion and for participants to develop their own views in relation to the range of standpoints. This resulted in constructive dialogue between the many different points of view. The ultimate outcome of the seminar was that participants of the two main approaches seem to have come to a greater awareness of the links between and characteristics, strengths and weaknesses of both approaches. The discussions brought into focus the different conceptualisations of literacy (whether single or multiple) as well as the choice of methods used to assess literacy.

The seminar discussions largely tended towards a dichotomy between basic skills and measurement of literacy skills per se; and qualitative approaches to assessment and the social view of literacies, although the ‘functional literacy’ approach also featured. As referred to earlier, recent work by Brandt and Clinton (2002) has stressed that an alternative approach can be found between the two approaches to literacy. However, participants argued that the different approaches to assessing literacy cannot be lined up as a continuum: although they are not mutually exclusive as research orientations, the different perspectives and their associated concepts, techniques and methods cannot be treated as complementary (Leung, from comments). Furthermore, ethnography can contribute to assessment without being subsumed into the measurement approach (Kell, from comments).

Literacy assessment needs to account for the range of motivations for learning literacies. Participants may want to master literacy competencies for their own (individual) benefit or interest, whether this is personal, religious or work-based. In some cases they may see assessment of literacy competencies as a stepping-stone to more formal qualifications in the future (see Kell 1996) and thus be more interested in the technical and linguistic aspects of literacy. However, in other cases, participants may be in need of and more interested in equity, ‘development’ and group outcomes for themselves and their communities, and this is where assessment needs to take account of the social, multiple approach to literacies (as in CLPN). Ethnographic work can contribute to the assessment process through providing such data.

Given the current dominant influence of cost effectiveness and accountability considerations, distinct types of assessment are likely to be required to satisfy the differences in scale, agencies and actors involved. This does not mean that ethnography must be forced to translate its findings into quantifiable and measurable results. Despite differences in underlying assumptions (Teasdale and Leung 2000; Hamilton and Barton 2000), this seminar has shown that ethnographic and qualitative methods, and a multi-method approach, can contribute in key ways to the quality of quantitative literacy assessment procedures: some came to the view that it is “not either or but both” (Greaney, comment). It has also suggested that conventional, statistical methods to assessing literacy do not prohibit the incorporation of the multiple view

of literacy, and that the failure to do so may be a result of resistance to diversity (Freeland, from notes) rather than purely technical considerations.

**Future Directions:**

The seminar produced possibilities for moving forward in two main areas:

**The LAMP programme:**

A group formed around the LAMP programme with the following aims in the future:

- to carry out an ethnographic study of the LAMP process during the piloting stages.
- to contribute with ethnographic input into the programme itself, helping to define and clarify constructs and items used for assessment, to balance the statistical methods of LAMP and further refine the methodology used (Freeland; Imperial, from comments).

It was thought that LAMP may be useful in developing countries as an alternative to national surveys for the purposes of larger educational and social planning, if methodological concerns and philosophical underpinnings could be agreed by participating countries (Imperial, from comments).

**Multi-lingual literacies**

A group that formed and ran throughout the length of the seminar agreed to maintain contact in order to take some of the issues and discussions further with a view to organising a conference on this theme in the future (Freeland, from comments).

\*\*\*\*\*

## APPENDIX

In order to do justice to the process of the seminar, this section incorporates a more detailed account of the particular presentations given during the proceedings and of some of the papers circulated before the seminar began.

### 1. The Process of the Seminar:

The seminar began with a background paper written by Alan Rogers and Cathy Kell, and presented by Cathy Kell. Then a number of participants gave short presentations on their experience, including Bénédicte Terryn on the UNESCO LAMP programme, Constant Leung on psychometric testing and alternative assessment, Regie Stites on the Equipped for the Future (EFF) approach, Kara Jackson on literacies in multi-lingual environments, Aues Scek on participatory monitoring in Kenya, and Bryan Maddox and Roshan Chitrakar on a case study in Nepal. These were all based on papers sent out to all participants before the seminar. This was followed by a general discussion setting the agenda for the next day, with plenary and small group sessions. Plenary discussions during the three days were chaired by Brian Street, Mary Hamilton, Judy Kalman and Jane Freeland.

Three main questions and themes emerged from the initial discussion: ‘Why’, ‘what’ and ‘how’? The ‘why’ questions and discussion focussed on why we want to measure literacy and indeed if we want to do so, for what purpose and for whose benefit? A format for the ‘why’ questions was discussed to tease out different aspects of the issue in terms of social, economic and political elements. The ‘what’ and ‘how’ questions arose out of and were connected to the ‘why’ questions.

### Individual discussion papers and presentations:

The following descriptions are based on presentations and seminar papers for discussion circulated prior to the seminar:

#### Background paper (Kell and Rogers)

The background paper for discussion at the seminar was written by Cathy Kell and Alan Rogers. It raised some of the key critical questions for discussion around the assessment of literacy:

1. What is the purpose of measuring literacy: why and for whom?
  - to produce comparative national statistics or for national policy making;
  - to improve implementation of learning programmes;
  - to satisfy donors’ accountability requirements;
  - to give learners feedback on their learning.
2. Multiple literacies and power:
  - If multiple literacy practices are tied up with power relationships, is it dominant literacy practices that we are measuring using conventional methodologies, particularly at the national scale?
  - Do we therefore need different tools to measure multiple literacies?
  - Can using a view of multiple literacy practices lead to more valid and reliable ways of measuring literacy?
3. Scale
  - Are we measuring **national** literacy levels or **programme** levels of achievement? Can we use the same tools throughout?

4. Competencies or uses?
    - Are we measuring literacy **competencies** or **uses** of literacy, and if so, how?
  5. Measuring tools
    - Can we suggest tools for measuring literacy practices?
- (adapted from Kell and Rogers background paper)

These were all developed further during the ensuing discussions.

## 2. Models and Model Building – Brian Street (see above)

Some quotes from a social scientist (Weber 1949) and a mathematician (Gowers 2004).

### Weber on ideal models

We have in abstract economic theory an illustration of those synthetic constructs which have been designated as *ideas* of historical phenomena. It offers us an ideal picture of events on the commodity market under conditions of a society organized on the principles of an exchange economy, free competition and rigorously rational conduct. This conceptual pattern brings together certain relationships and events of historical life into a complex, which is conceived as an internally consistent system. Substantively, this construct in itself is like a *utopia* which has been arrived at by the analytical accentuation of certain elements of reality. Its relationship to the empirical data consists solely in the fact that where market-conditioned relationships of the type referred to by the abstract construct are discovered or suspected to exist in reality to some extent, we can make the *characteristic* features of this relationship pragmatically *clear* and *understandable* by reference to an *ideal-type*. This procedure can be indispensable for heuristic as well as expository purposes. The ideal typical concept will help to develop our skill in imputation in *research*: it is no ‘hypothesis’ but it offers guidance to the construction of hypotheses. It is not a *description* of reality but it aims to give unambiguous means of expression to such a description. It is thus the ‘idea’ of the *historically* given modern society, based on an exchange economy, which is developed for us by quite the same logical principles as are used in constructing the idea of the medieval ‘city economy’ as a ‘genetic’ concept. When we do this, we construct the concept ‘city economy’ not as an average of the economic structures actually existing in all the cities observed but as an *ideal-type*.

An ideal type is formed by the one-sided *accentuation* of one or more points of view and by the synthesis of a great many diffuse, discrete, more or less present and occasionally absent *concrete individual* phenomena, which are arranged according to those one-sidedly emphasized viewpoints into a unified *analytical* construct (*Gedankenbild*). In its conceptual purity, this mental construct (*Gedankenbild*) cannot be found empirically anywhere in reality. It is a *utopia*. Historical research faces the task of determining in each individual case, the extent to which this ideal-construct approximates to or diverges from reality, to what extent for example, the economic structure of a certain city is to be classified as a ‘city-economy’. When carefully applied, those concepts are particularly useful in research and exposition. In very much the same way one can work the ‘idea’ of ‘handicraft’ into a utopia by arranging certain traits, actually found in an unclear, confused state in the industrial enterprises of the most diverse epochs and countries, into a consistent ideal-construct by an accentuation of their essential tendencies. This ideal-type is then related to the idea (*Gedankenausdruck*) which one finds expressed there. One can further delineate a society in which all branches of economic and even intellectual activity are governed by maxims which appear to be applications of the same principle which characterizes the ideal-typical ‘handicraft’ system. Furthermore, one can juxtapose alongside the ideal typical ‘handicraft’ system the antithesis of a correspondingly ideal-typical capitalistic productive system, which has been abstracted out of certain features of modern large scale industry. On the basis of this, one can delineate

the utopia of a 'capitalistic' culture, i.e., one in which the governing principle is the investment of private capital.

This procedure would accentuate certain individual concretely diverse traits of modern material and intellectual culture in its unique aspects into an ideal construct which from our point of view would be completely self-consistent. This would then be the delineation of an 'idea' of *capitalistic culture*. We must disregard for the moment whether and how this procedure could be carried out. It is possible, or rather, it must be accepted as certain that numerous, indeed a very great many, utopias of this sort can be worked out, of which *none* is like another, and *none* of which can be observed in empirical reality as an actually existing economic system, but *each* of which however claims that it is a representation of the 'idea' of capitalistic culture. *Each* of these can claim to be a representation of the 'idea' of capitalistic culture to the extent that it has really taken certain traits, meaningful in their essential features, from the empirical reality of our culture and brought them together into a unified ideal-construct. For those phenomena which interest us as cultural phenomena are interesting to us with respect to very different kinds of evaluative ideas to which we relate them. Inasmuch as the 'points of view' from which they can become significant for us are very diverse, the most varied criteria can be applied to the selection of the traits which are to enter into the construction of an ideal-typical view of a particular culture.

What is the significance of such ideal-typical constructs for an *empirical* science, as we wish to constitute it? Before going any further, we should emphasize that the idea of an ethical *imperative*, of a 'model' of what 'ought' to exist is to be carefully distinguished from the analytical construct, which is 'ideal' in the strictly logical sense of the term. It is a matter here of constructing relationships which our imagination accepts as plausibly motivated and hence as 'objectively possible' and which appear as *adequate* from the nomological standpoint.

*Weber 1949:89-93*

### **Gower on what is a mathematical model?**

When one examines the solution to a physical problem, it is often, though not always, possible to draw a clear distinction between the contributions made by science and those made by mathematics. Scientists devise a theory, based partly on the results of observations and experiments, and partly on more general considerations such as simplicity and explanatory power. Mathematicians, or scientists doing mathematics, then investigate the purely logical consequences of the theory. Sometimes these are the results of routine calculations that predict exactly the sorts of phenomena the theory was designed to explain, but occasionally the predictions of a theory can be quite unexpected. If these are later confirmed by experiment, then one has impressive evidence in favour of the theory.

The notion of confirming a scientific prediction is, however, somewhat problematic, because of the need for simplifications of the kind I have been discussing. To take another example, Newton's laws of motion and gravity imply that if you drop two objects from the same height then they will hit the ground (if it is level) at the same time.

This phenomenon, first pointed out by Galileo, is somewhat counter-intuitive. In fact, it is worse than counter-intuitive: If you try it for yourself, with, say, a golf ball and a table-tennis ball, you will find that the golf ball lands first. So in what sense was Galileo correct?

It is, of course, because of air resistance that we do not regard this little experiment as a refutation of Galileo's theory: experience shows that the theory works well when air resistance is small. If you find it too convenient to let air resistance come to the rescue every time the predications of Newtonian mechanics are mistaken, then your faith in science, and your admiration for Galileo, will be restored if you get the chance to watch a feather fall in a vacuum - it really does just drop as a stone would.

Mathematicians do not apply scientific theories directly to the world but rather to *models*. A model in this sense can be thought of as an imaginary, simplified version of the part of the world being studied, one in which exact calculations are possible.

*from Gowers 2004*

### **3. Measurement approaches: seminar papers and presentations**

Several of the participants put forward versions of and arguments for classical measurement approaches. This included Item Response Theory, Psychometric testing and the Literacy Assessment and Monitoring Programme (LAMP) devised by the UNESCO Institute of Statistics (UIS) in partnership with other agencies such as the World Bank. A more qualitative approach was also put forward, such as in the KePIM programme.

Measurement of literacy is not new, but interest in it has increased as a result of international conferences and initiatives on literacy over the last two decades such as the Education for All conferences at Jomtien and Dakar. Pressure has increased for accurate measurement and the production of comparable statistics across countries in order to assess progress towards the Jomtien and Dakar goals. The conferences have also brought into sharper focus the inadequacies of previous methods of measuring literacy (UNESCO 2002). Many large-scale literacy assessments are based on census (of whole population) or household survey data (using sample approach), sometimes supplemented or replaced by dedicated literacy surveys. Classical-type assessments test only 'prose' (UIS leaflet) or 'essay text' literacy (Kell 2003) and basic numeracy. They usually take place at fixed time intervals e.g. ten years for census data, and thus provide a 'snapshot' view.

#### **a) LAMP - Benedicte Terryn**

LAMP is a new programme intended to improve on previous methods of measuring literacy for the purpose of comparison across countries. The LAMP programme evolved out of the International Adult Literacy Survey (IALS), an influential international assessment programme, impacting on many countries' implementation of literacy programmes, including in the UK and Ireland (Crowther, Hamilton and Tett 2001; Merrifield, seminar paper). The IALS showed that the lack of certain literacy skills was more widespread than many people had thought, particularly in 'developed' countries.

LAMP is currently being piloted in a few countries; it uses a method based on Item Response Theory (see below) and a sample approach, which can be projected to the country level, and is considered an improved and refined version of previous large-scale surveys aimed at measuring literacy competencies (Stites, from notes on small group discussion). It includes additional questions on background data of respondents to highlight some contextual aspects. LAMP, as in other large-scale literacy surveys, seeks to compare data across countries and over time.

This process of assessment incorporates extra 'levels' of literacy. LAMP intends to test literacy at a range of levels of competence, using either a profiling or two-level approach to distinguish those that can perform 'basic' literacy tasks and above from those who are unable to read and write at all. LAMP is designed to test mainly 'prose' although it attempts to take into account local realities (UIS leaflet). The surveys are to be carried out in partnership with and promoting the capacity of national governments to carry out surveys of this type. National governments will decide, for example, which languages will be included in the literacy survey.

However, a number of criticisms were raised by participants in relation to LAMP:

At such a large scale, considerations of efficiency in cost and time are crucial, as are summary measures; the methodology can only be responsive to local contexts in a limited way.

Devolution of decision-making about the language(s) of assessment to individual countries themselves has been criticised. This was challenged by a number of participants on political grounds as it assumes that governments always work in the interest of all of their people regardless of social and power relations and the interests represented by dominant language practices in that country.

Many countries in the ‘South’ may be unable to keep up with the rapid theoretical and methodological changes in literacy assessment over recent years, including the change from IALS to LAMP, and the extra resources this requires from national governments. Changes in methodology also call into question the validity of making comparisons between surveys of a country over time.

LAMP has also been criticised for failing to incorporate literacies in other languages than the dominant one(s) of the country, and its definition seems to question the usefulness of measuring localised languages (Jackson, seminar paper):

“Literacy involves a continuum of learning to enable an individual to achieve his or her goals, to develop his or her knowledge and potential, and to *participate fully in the wider society*” (UIS 2003:4, emphasis added).

There are social and political implications of classifying people as having a ‘low level’ of literacy in the dominant language when they may have a ‘high level’ in one or more local languages.

One of the household survey approaches used for literacy incorporates a clause that automatically ascribes a high level of literacy when the respondent is literate in more than one ‘significant’ language. The household survey approach is the most commonly favoured internationally; it incorporates multiple levels rather than just literate/non-literate, it uses a sample approach, and is less expensive than IALS and its subsequent survey Adult Literacy and Lifeskills survey (ALL) (Jackson, seminar paper).

*Incorporating ethnography into LAMP:* Precedents exist for using ethnography at the preliminary planning stages of literacy implementation such as in the Community Literacy Project Nepal. However, in addition to time and resource constraints, political constraints may act to resist the carrying out of ethnographic work. National governments may feel they know what literacies exist in their country and therefore may not see the necessity for an ethnographic component prior to the actual literacy survey (Jackson, seminar paper).

## **b) Item Response Theory and Psychometric testing – Constant Leung**

Notes in connection with a paper were circulated for the seminar (Teasdale and Leung 2000) on school education and English as a Second Language (ESL) in primary schooling. Constant Leung’s paper and presentation gave a more in-depth insight into the relevance and issues concerning psychometric testing and Item Response Theory in relation to the different approaches to understanding and quantifying literacy / literacies. The issues covered included some of the problems with the concept of validity and with attempts at using alternative or performance based assessment.

IRT is used in standardised or ‘classical’ testing, assessing in a ‘controlled’ environment using a set of questions (items). Standardised or classical testing and IRT approaches to analysis are associated with psychological and psychometric theories of testing. Psychometric

research measures characteristics, skills, abilities or attitudes of *individuals* (or ‘constructs’), which are derived from theory (involving abstraction) and operationalised as items (questions) in a test. Examples of constructs in education are intelligence, or reading ability. Psychometric research is concerned with the reliability of measuring constructs, which must be lasting characteristics (from unpublished paper written specifically for seminar).

Psychometric theory makes a number of assumptions. Items measure individual psychological characteristics and do not take into account group work and achievements in ‘real time’. Psychometric theory assumes that we can pre-specify what to test and then devise items: that testing in this way produces valid results. Items (questions) have only correct/incorrect answers (they are scored dichotomously).

Constructs (abilities) are assumed to be uni-dimensional; each item in a test measures a single underlying attribute. This means that it is incompatible with the multiple, social view of literacies. Items are excluded if they do not correlate highly with the total score because a test (or one part of a test) is concerned with measuring only one attribute, for example, ‘reading’ often defined in terms of single attribute of accuracy or fluency. (If there is more than one psychological process involved, it is assumed that they function in unison.)

It was pointed out that psychometric approaches to assessing literacy are therefore *incompatible* with the *multiple*, social view of literacies, since they can only measure *one-dimensional* constructs (Stites, from comments).

Psychometric theory assumes that the relationship between observable performance and underlying abilities can be determined. In IRT, a person’s ability is conceptualised as ‘latent variables’ underlying behaviour; responses on a test provide an indication of a person’s ability on the latent variable<sup>8</sup>.

IRT assumes that we can relate ‘ability’ (the respondent’s location on the scale, corresponding to the score) to ‘difficulty’ (the location of the test on the scale) using the same rating scale. IRT also assumes that subjects are in bounded domains, not overlapping. This is not the case in assessment of many learning programmes, and presents particular problems in multi-lingual environments where specific language and cultural knowledge is often implicit and cuts across all competencies.

With alternative assessment approaches, e.g. portfolio or ‘real’ tasks and ‘real time’ interaction, sets of rating scales can be used. The difficulty is how to validate these in a non-controlled environment since scores are based on teacher judgement. Therefore it is not possible to validate them by using procedures employed in psychometric testing. This kind of assessment is formative, not summative.

Teasdale and Leung’s paper (2000) makes a number of points to do with validity in alternative approaches. They suggest that insufficient investigation has been carried out into how elements of assessment can be incorporated into the learning and teaching process, and how classroom-based teacher assessment can be understood as a form of formative assessment<sup>9</sup>. They emphasise the potential danger in alternative approaches of not attending

---

<sup>8</sup> Representation: literacy assessment is concerned with representing (placing on a rating scale) a representation (the score) of the particular literacy/numeracy competence; this is the ‘latent variable’; literacy assessment is therefore a second order representation.

<sup>9</sup> Teasdale and Leung (2000) list four complementary purposes for assessment. The first three are for assessment of learners, the last is at institutional level: formative (assessing positive achievements and where to go next); diagnostic (of areas needing attention); summative (systematic summary of overall achievements); and evaluative (educational institution).

to the different meanings and pedagogic values of formal standardised testing and teacher assessment. Boundaries between testing and teacher assessment may be blurred, and the meaning and validity of scores produced by teacher assessment taken as unproblematic (assuming that psychometrically oriented principles can be applied). The participant role of the teacher and the lack of stability of tasks present problems for validation using psychometric principles. One of the reasons given for this is the different roles and functions that the teacher carries out in any learning situation: a regulatory function which maintains social hierarchical factors (plus discipline and rules such as turn-taking to enable learning to take place), and a management function in charge of the content and learning processes. The influence of the teacher and potential sources of bias are not taken into account in performance/teacher assessment.

The conceptualisation of validity is one of the crucial distinctions perceived between psychometric and alternative approaches to testing. Teasdale and Leung (2000) argue that the two approaches have different concepts of validity; that alternative approaches to assessment base claims of validity on the resemblance with 'real life' tasks and on a procedure rather than on psychometric properties of test items. Advocates of alternative approaches seem to borrow from validation used by psychometric testing whilst at the same time distancing themselves from psychometric approaches (Teasdale and Leung 2000). It is argued that there is, at the present time, a lack of development in the conceptualisation of the notion of validity in alternative assessment; real-life tasks are assumed to be valid because they are embedded in everyday reality of the students. On this view validity is conceptualised as a property of procedure (rather than in test scores as in psychometric testing). Thus there is a lack of clarity in the way validity is claimed.

At the same time Leung pointed out that psychometric theory, as it is currently understood, is unable to accommodate itself to student- and learning-sensitive approaches to curriculum and pedagogic concerns. It is incompatible with values underlying particular pedagogies and epistemologies such as constructivism. The high cost of test development may also inhibit full implementation in certain educational systems.

### **c) Equipped for the Future (EFF) – Regie Stites**

This paper outlined an approach to adult basic education and literacy that combines conventional and alternative approaches to measurement. It is based on a set of standards used to assess outcomes for the purposes of accountability and evaluation at the programme level, aimed at improving the quality of the system. Generally, standards for assessment of programmes have in the past been poorly aligned with accountability requirements. The EFF approach seeks to address this problem of poor compatibility and to clarify how standards of assessment and accountability can enhance educational practice and lead to higher levels of learning.

*Alternative assessment:* EFF incorporates alternative, evidence-based methods of assessment linked to standards. The EFF approach stresses the importance of dissemination of information and knowledge on assessment and best practice at all levels throughout the whole system including to teachers and to the learners themselves. Assessment is performance-based (oral and written), and aligned as far as possible with 'real life' tasks. Recent studies including in cognitive science have found a link between learner motivation and relevance, showing how closely assessments are connected to learning and life goals and learners' everyday lives. Research has shown that adult learners are often unmotivated or resistant to standardised test formats.

Alternative or performance assessment means that learners can have a greater degree of determination of the learning goals. A degree of flexibility is possible in learning programmes

and context can be taken into account. Expressed goals in specific contexts can be incorporated (e.g. English language abilities for asking for directions around town).

Validity of performance assessment can be achieved by clearly specifying the content domain (or construct) to be measured and structuring tasks to ensure the production of sufficient evidence to support assessment interpretations. Fairness and accuracy of testing using alternative methodologies are also critical considerations. The EFF approach seeks to ensure this by examining carefully consequences of test use and interpretation and correcting features of assessments that may bias results and interpretations in favour of one group of learners over another. The EFF approach therefore addresses some of the concerns raised above about alternative forms of assessment.

A key issue was reinforced by the presentation on the EFF approach, that often discussions about assessment can become framed in terms of what should be measured and how, to the exclusion of why: for what purpose and for whose benefit? The 'why' questions were later taken up in the plenary discussion.

#### **d) Processes of Measuring Literacy and Numeracy Attainment – Dave Baker**

The presentation on numeracy made clear something that is often forgotten about measurement; that measurement and numerical calculations often involve value judgements, they are rarely neutral or value-free. The following account addresses some of the concerns about how we measure literacy and numeracy, and about the dangers of representing complex practices like literacy and numeracy by one or more numbers.

During the process of numerical calculations, characteristics in a particular social context are given as a number, thus simplifying them into a model and taking them away from their context and meaning. Leaving out detail may make the model invalid if it affects attributes that need to be preserved such as order and linearity. Calculations are then carried out on the numbers as if they did not affect the original situation in a significant way. To manage the data, indicators are often chosen, using a culturally produced value system; such choices need to be acknowledged and justified.

Calculations are more or less appropriate and meaningful according to the context and type of the original numbers (such as whether they are ratio data). Heights can be averaged meaningfully, but not rankings as they are often not comparing like with like (such as averaging rankings of aspects of educational performance to produce an overall score, then comparing with other scores).

After the calculations are carried out, numbers are mapped back into the original context, reconnected to their social relations and reinterpreted. They are then used to influence policy. Dominant practice often does not make visible or justify judgements and choices made during this process. Numbers can gain an unjustified authority because of their surface simplicity, as well as an unjustified accuracy due to the precision in which results are presented (Dave Baker, seminar paper: see diagram).

Given that this is the case, the fact that numeracy is often presented as the poor cousin to literacy is surprising (from presentation). This includes the ways that numeracy is seen and even is assessed and measured in education and especially in adult education. Too frequently numeracy is seen as a part of literacy rather than as a valuable and valid area in its own right

### **e) Multi-lingual literacies – Kara Jackson: A Case Study in Vanuatu**

In the Vanuatu case study, Jackson recommended carrying out a preliminary investigation into what languages are used on a regional basis within the country and which are written down (e.g. history of formal script) before carrying out a large-scale literacy survey. She emphasised the consequences of carrying out such a literacy survey. Choices made about which languages and practices are counted as literacies will valorise particular languages and literacies over others in that country; this shows the ‘ideological’ nature of literacies (Street 1993).

Jackson also highlighted the problems that can occur when stakeholders in a literacy survey may not agree on the purpose of carrying out the survey. They may have hugely differing aims such as to privilege one (non-colonial) language over others, to produce an overarching percentage of the literacy rate, or to inform broad educational policy.

In addition, Jackson discussed the concept of ‘levels’ of literacy and asserted that literacy viewed as a set of ‘levels’ can be misleading as has been shown by some of the IALS results (UIS leaflet).

### **f) Ethnographic/Qualitative approaches to Impact Monitoring: The Kenya Participatory Impact Monitoring programme (KePIM) – Aues Scek**

Scek presented the papers on KePIM which he tabled at the seminar. KePIM is linked directly to poverty reduction: it is integrated with the monitoring and evaluation system for the Poverty Reduction Strategy Policy (PRSP). Basic education was one of the four key areas identified for action through previous participatory poverty surveys (Wainaina 2003). The programme was designed to inform policy process directly, and incorporated participatory monitoring processes including PRA (participatory rural/rapid appraisal) as well as other methods.

KePIM replicated a previous qualitative monitoring process and report carried out in Malawi. Aues Scek outlined how a large-scale national survey of educational programmes in Kenya including enrolment in school education was carried out. What they found was low school enrolment, but insufficient information to find out why enrolment was low, therefore a participatory research project was set up. Through a qualitative participatory methodology, they found that the problem was largely caused by families not being able to afford school fees and were able to rectify this by stopping the charges. In this way the qualitative phase of the assessment process was able to enhance understanding and get at motivations in a way that the national survey could not have.

### **g) Literacies and post-literacy in Nepal: the Community Literacy Project Nepal (CLPN) - Bryan Maddox and Roshan Chitrakar**

The Community Literacy Project Nepal has been working with groups interested in local development issues. It was conceived as a literacy project working within the New Literacy Studies model. The project focussed on group participation and the mediating of literacy within the group to achieve complex literate tasks. The tasks were issues of importance to the group and local community, including related to accessing forest resources. The main concern was not making each individual literate or testing literacy competences.

Three different spheres of activity took place (not necessarily a linear progression):

- a) Institutional change - group activities/achievements:
  - Social justice movement
  - Translation of forest constitution
  
- b) Collective practice - specific activities carried out by group:
  - Filling in citizenship form
  - Completing death/marriage certificate
  - Letter writing
  - Group minute writing
  - Group savings calculations
  
- c) Additional literacy tasks carried out by some individuals
  - Writing/signing name
  - Keeping written records
  - Making simple calculations

*Mediation of literacy:* The CLPN programme illustrates the mediation of literacy and numeracy practices in the group to achieve the specific collective practice and institutional change aims. Literacy mediation, where people are able to access other community members' skills and knowledge of particular literacy contexts and practices such as official literacies, has been documented elsewhere (such as Baynham 1993; Kell 2003). Thus people who may lack particular specific skills in, for example, form-filling literacy in official contexts are fully able to participate in their literate environment (Baynham 1993; Clinton Robinson, from notes).

*Questioning measurement and scales:* The project found that people's 'competencies' in literacy didn't affect their ability to participate in the group process. Literacy competencies couldn't necessarily be mapped onto linear scales; some persons were able to engage at a 'higher level' without the 'lower level' ability.

*Pressure to produce measurable statistics of literacy programmes:* The outcomes of the project were not helpful to national literacy statistics since individual competencies and improvements could not be measured. The project therefore lost funding as literacy outcomes were group rather than individual learning outcomes. However, the project was later able to secure funding as a 'post-literacy' initiative.

During the discussion on the CLPN programme, Vincent Greaney distinguished between second-order outcomes of literacy intervention (such as the above outcomes relating to 'participation', 'community development' and 'empowerment'), and first order outcomes such as individual literacy competencies.

The discussion around the CLPN programme seemed to produce a degree of polarisation of views into those who saw the aims of literacy programmes as promoting 'community development' and 'empowerment', social justice and equity, those who focused on the 'uses' of literacy, whether related to 'development' or not, and those who were concerned more with the measurement of literacy competencies per se rather than the uses of the competencies. The different standpoints and disciplines of the participants created a space for constructive dialogue. Assumptions had been voiced earlier that ethnography had to translate its findings into quantitative form; this was now challenged:

'From the perspective of my particular interest in multilingualism, this seemed analogous to the demand that people who speak the minority language should

become bilingual, whereas people who speak the dominant language do not need to. Neither language is intrinsically superior, but historical processes have made one of them dominant and hegemonic' (Jane Freeland, her notes on the multi-lingual literacies group).

Discussion on the third and last day, however, seemed to be marked by a greater awareness by participants of the characteristics, strengths and weaknesses of other approaches, and an appreciation of the contributions that ethnographic and qualitative approaches could make to the more formal assessment processes. A greater acceptance of common ground was evident.

## **Uppingham Seminar 27-29<sup>th</sup> November 2003**

### **Timetable**

#### **Thursday**

Session 1: 4pm – 6.30pm

a) Background paper (Rogers and Kell); presented by Cathy Kell

Presentations included: Regie Stites (EFF); Bénédicte Terryn (LAMP); Constant Leung (psychometric and alternative forms of testing); Kara Jackson on multi-lingual literacies in Vanuatu; Aues Scek on KePIM in Kenya; Dave Baker on numeracy

b) Discussion chaired by Brian Street

Why, what, how questions emerged.

7pm Thanksgiving Dinner

#### **Friday**

Session 2: 9am - 1pm: chaired by Mary Hamilton

a) Plenary session: Why questions.

b) Groups: Discussing why – what questions at micro level and large scale considering social, political and economic aspects.

1pm – 2.30pm: Lunch

Session 3: 2.30pm – 6.15pm

Plenary chaired by Judy Kalman

‘How-ish’ questions (included in the appendix):

Short presentations on various projects related to the questions raised including: Aues Scek on KePIM in Kenya, India, Regie on EFF in USA; Bénédicte on LAMP (UIS, UNESCO); Bryan Maddox and Roshan Chitrakar on CLP, Nepal and further discussion.

7.30pm onwards: Dinner and live jazz music.

#### **Saturday**

Session 4: 9am-1.30pm

a) Plenary chaired by Jane Freeland.

b) Groups: one group discussed multi-lingual literacies, multiple literacies and scripts; the other three discussed issues connected with ethnography and measurement at the micro-level and large scale (see notes above).

1.30pm – 3.15pm: Lunch and depart

## References

- Baker, D. (2003). 'Processes of Measuring: Literacy and Numeracy attainment', *Uppingham Seminar Paper*.
- Barton, D. and Hamilton, M. (1998). *Local Literacies. Reading and Writing in One Community*, London: Routledge
- Baynham, M. ed (2004). 'Futures of the ethnography of literacy?' in *Language in Education Special Issue on Ethnographies of Literacy*.
- Baynham, M. (1993). 'Code switching and mode switching: Community interpreters and mediators of literacy' in Street, B.(ed) (1993). *Cross-cultural approaches to literacy*, Cambridge: Cambridge University Press:294-314).
- Bernstein, H. et al. (1992). *Rural Livelihoods: Crises and Responses*, Oxford: Oxford University Press/Open University.
- Brandt, D. and Clinton, K. (2002). 'Limits of the Local: Expanding Perspectives on Literacy as a Social Practice' *Journal of Literacy Research* 34(3):337-356.
- Crowther, J., Hamilton, M. and Tett, L. (2001). 'Powerful Literacies: an introduction' ch.1 in *Powerful Literacies*, Leicester: National Institute of Adult Continuing Education (NIACE):1-12.
- Easton, P. (no date) 'Defining and measuring literacy' Annex 1, World Bank:21-24
- EFF (forthcoming) *Reporting Achievement, Improving Results. The EFF Approach to Accountability: An Assessment Prototype for Read With Understanding*, draft document, Equipped for the Future Assessment Consortium, USA.
- Gee, J.P. (1990). *Social Linguistics and Literacies: Ideology in Discourses*, London: The Falmer Press
- Giddens, A. (1997). *Sociology*, third edition, Cambridge: Polity Press
- Goffman, E. (1974). *Frame Analysis: An Essay on the Organisation of Experience* New York:Harper and Row,
- Gowers, T. (2004). *Mathematics: a very short history*, Oxford: Oxford University Press
- GTZ Social Policy Advisory Services (2003). *Kenya Participatory Impact Monitoring Systems - (KEPIM)*, Information for Uppingham Seminar, available on <http://www.gtzspas.com/kepim.htm> [downloaded Jan.04].
- Hamilton, M. and Barton, D. (2000). 'The International Adult Literacy Survey: What Does It Really Measure?' *International Review of Education*, 46(5): 377-389
- Herbert, P. and Robinson, C. (2001). 'Another Language, Another Literacy? Practices in Northern Ghana' in Street, B. (ed.) (2001) *Literacy and Development: Ethnographic Perspectives*, London: Routledge:121-136).  
*International Review of Education* Vol.46 (5) (2000) Special edition.
- Jackson, K. (2003). 'Literacies and Development: The Implications of an NLS Perspective on Literacy Survey Design: A Case Study in Vanuatu', *Uppingham Seminar Paper*.
- Kell, C. (1996). 'Literacy practices in an informal settlement in the Cape Peninsula in Prinsloo, M. and Breier, M. (eds.) (1996) *The Social Uses of Literacy: Theory and Practice in Contemporary South Africa*, South Africa: Sached Books and Amsterdam: John Benjamins Publishing,
- Kell, C. (2003). *Accounting for Not Counting: Ethnography and Literacy in South Africa*, Paper presented at National Research and Development Centre Conference, 'What Counts for What Purposes in Research in Adult Literacy, Numeracy and ESOL?' Nottingham, April 2003; available on [http://www.uppinghamseminars.org/literacy\\_matters.htm](http://www.uppinghamseminars.org/literacy_matters.htm) (downloaded December 2003).
- Kell, C. and Rogers, A. 'Background paper', *Uppingham Seminar Paper*.
- Leung, C. (2003). 'Notes on standardised testing and Item Response Theory (IRT)', *Uppingham Seminar Paper*.
- Merrifield, J. (2003). 'Literacy assessment in Ireland: the NALA Assessment Framework', *Uppingham Seminar Paper*.

- Puchner, L. (2003). 'Women and literacy in rural Mali: a study of the socio-economic impact of participating in literacy programs in four villages', *International Journal of Educational Development* 23 (2003):439-458, available online at [www.elsevier.com/locate/ijedudev](http://www.elsevier.com/locate/ijedudev).
- Robinson, C. (2001). *Literacy and the White Paper 'Eliminating World Poverty: making globalisation work for the poor'*, paper presented at the 'Literacies and Livelihoods Seminar' organised by DFID, 28 June 2001.
- Rogers, A. (1999). 'Improving the quality of adult literacy programmes in developing countries: the 'real literacies' approach' in *International Journal of Educational Development* 19:219-234.
- Rogers, A. (2000). 'Literacy comes second: working with groups in developing societies' in *Development in Practice*, 10(2):236-239.
- Rogers A (2001) Afterword: problematising literacy and development, in Street (2001)
- Stites, R. (2003). 'Assessing Results that Matter: Equipped for the Future's Approach to Assessment for Adult Basic Education Accountability and Improvement', EFF Assessment Consortium, *Uppingham Seminar Paper*.
- Street, B.(ed.) (1993). *Cross-cultural approaches to literacy*, Cambridge: Cambridge University Press.
- Street, B. (1994). 'Cross-cultural approaches to literacy' in Maybin 1994.
- Maybin, J. (ed.) (1994). *Language and literacy in social practice : a reader*, Clevedon: Multilingual Matters in association with the Open University
- Street, B. (1996). 'Review of International Adult Literacy Survey', in *Literacy across the Curriculum*, Centre for Literacy, Montreal vol. 12, no. 3, Winter 1996:8-15 (reprinted 1998 as CfL Occasional Paper No. 1, with Graff, Jones etc).
- Street, B. (ed.) (2001). *Literacy and Development: Ethnographic Perspectives*, London: Routledge
- Street, B. (2003). 'What's 'new' in New Literacy Studies? Critical approaches to literacy in theory and practice' in *Current Issues in Comparative Education*. 5(2) May 12 available on <http://www.tc.columbia.edu/cice/>
- Teasdale, A. and Leung, C. (2000). 'Teacher assessment and psychometric theory: a case of paradigm crossing?' *Language Testing* 17(2):163-184.
- Terryn, B. (2003). The Literacy Assessment and Monitoring Programme (LAMP): Background and Approach, UNESCO Institute for Statistics, *Uppingham Seminar Paper*.
- UNESCO (2002) Global Monitoring Report: *Is the World on Track?* Paris: UNESCO
- UNESCO Institute for Statistics (UIS). (no date). Literacy assessment and monitoring programme: LAMP. [Brochure]. Montreal: UIS.
- UNESCO Institute for Statistics (UIS). (2003). *Literacy assessment and monitoring programme (LAMP): A new approach to literacy measurement*, UIS, Montreal.
- Wainaina, S. (2003). *An Introduction to The Kenya Participatory Impact Monitoring Exercise (KePIM)*. Presentation by Stephen Wainaina, Head, Human Resources and Social Services Department, Ministry of Finance and Planning, Nairobi  
<http://www.worldbank.org/wbi/attackingpoverty/activities/dakar-wainaina.pdf>
- Weber, M (1949). *The Methodology of the Social Sciences*, translated and edited by Edward A. Shils and Henry A. Finch, New York: The Free Press