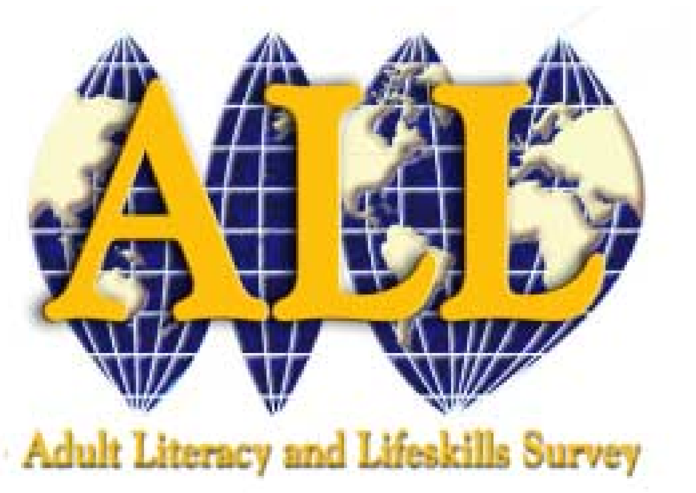
ALL
Adult Literacy and Lifeskills Survey

Numeracy

Numeracy

Executive Summary

Mathematical competency is one of the most widely assessed on both the national and international levels. Several studies, including the Third International Mathematics and Science Study (TIMSS) and the Program for International Student Assessment (PISA), have been, or will be, conducted to gain comparative international data on mathematics competence. One of the scales of the International Adult Literacy Survey (IALS), the Quantitative Literacy Scale, was a measurement of the respondent's ability to apply arithmetic operations to numbers embedded in diverse texts. While this scale produced useful data, survey developers recognized that it was limited in scope, as it was bounded by the use of text, among other things. The Numeracy scale of ILSS is designed to go above and beyond the QL Scale, while avoiding reliance on formal, curriculum-based knowledge of mathematics such as in TIMSS and PISA.

The conception of Numeracy utilized by ILSS is built not only upon the IALS conception of Quantitative Literacy, but also upon recent research and work done in the Netherlands, Australia, the United States, and other countries. The framework advances a definition of Numeracy as "The knowledge and skills required to effectively manage the mathematical demands of diverse situations".

As a full assessment of all elements of such a broad definition is beyond the scope of ILSS, the framework sets out to assess particular aspects of numerate behavior. Numerate behavior is divided into five key facets, each with subcategories: (1) managing a situation or a problem in a realistic context such as everyday life, work-related, societal or community, and further learning; (2) responding to a situation through identifying, interpreting, acting upon, and communicating; (3) information about mathematical ideas such as quantity & number, dimension & shape, pattern and relationships, data & chance, and change; (4) representations of mathematical information including objects and pictures, numbers and symbols, formulae, diagrams and maps, graphs, tables, and texts; (5) enabling knowledge bases and reasoning processes including contextual knowledge, mathematical (and statistical) knowledge and understanding, mathematical problem-solving skills, literacy skills, and beliefs and attitudes. In addition, a scheme of five factors was developed to account for the difficulty of different tasks, enabling an explanation of observed performance in terms of underlying cognitive factors.

Based on the conceptual framework developed, a pool of 80 items was developed and tested in a feasibility study in the U.S. and the Netherlands in June 1999. A report of the results of this study is enclosed in Appendix 1, with sample items in Appendix 2. Overall, the feasibility study has shown that the item pool includes tasks at diverse levels of difficulty and that cover key facets of the conceptual framework for numeracy. The theoretical factors that are supposed to account for task difficulty were found to be highly correlated with actual difficulty of items. Preliminary results provide initial support for the content validity and the construct validity of the numeracy scale. Suggestions for completion of item development and a tentative timeline are described.

Numeracy

This paper presents a framework for the conceptualization and assessment of numeracy skills as part of the International Life Skill Survey (ILSS). The beginning of this paper outlines the purpose of numeracy assessment, examines how numeracy is related to literacy and to school mathematics, reviews some factors known to affect adults' use of mathematical skills in real-world contexts, and discusses how numeracy has been defined and assessed in the past. Overall, this background discussion aims to help the reader understand why there was seen a need to go beyond the Quantitative Literacy scale used in the International Adult Literacy Survey (IALS) and assess a broader domain called "Numeracy" in the ILSS, and in what ways a test of numeracy may be different than a test of Quantitative Literacy or a test of mathematical skills given to school students.

Next, the paper presents a working definition of numeracy, elaborates on the facets of numerate behavior that are to be captured by items in the Numeracy scale, and specifies factors that influence the complexity or difficulty of numeracy tasks. Understanding of these facets and factors is a key to interpreting results pertaining to skill distribution in the populations to be studied. A subsequent section describes the development of the initial item pool and scoring guidelines, followed by details on test administration issues and the use of calculators and other tools as part of the testing process. Closing the discussion of the proposed assessment scheme is a section on the importance of assessing certain beliefs and practices that people may have, that impact on or are related to their numeracy, and explains why these should be studied by specific questions proposed for inclusion in the *background* questionnaire.

Based on the conceptual framework and pragmatic considerations described in the sections outlined so far, a pool of 80 items was developed and tested in a feasibility study conducted in two countries in June 1999. Details regarding the design and results of this study are reported in Appendix 1. This document ends with a summary of current status and next steps to be taken for successful completion of the development of the numeracy scale.

Why Include Numeracy in ILSS?

Numeracy is becoming increasingly important in diverse adult education and workplace sectors, following its apparent low profile for many years in published reports describing adult literacy and basic education programs. As workplaces are becoming more and more concerned with involving all workers in improving workplace efficiency and quality processes, the importance of numeracy skills is growing, and they have been shown to be a key factor in workplace success (Jones, 1995; Murnane, Willett & Levy, 1995).

The concept of numeracy is also related to the dialogue about the goals and impact of mathematics education in schools. More mathematics educators now encourage links between the knowledge and skills gained in the mathematics classroom and students' ability to handle real-life situations that require activation of mathematical knowledge and skills (NCTM, 1989; Heuvel-Panhuizen & Gravemeijer, 1991; Willis, 1990). Thus, information about adults' numeracy can be of value to educators operating at various levels. Finally, in a society in which the media constantly present information in numerical or graphical form, the ability to interpret quantitative and statistical messages is now seen as vital for all adults. Given the increasing need for adults to continuously adapt to changing citizenship and workplace demands (European Commission's White Paper, 1996), it is vital that nations have information about their workers' and citizens' numeracy in order to plan effective lifelong learning opportunities.

Challenges

While numeracy is viewed as a vital skill area, its conceptual boundaries and methods for assessing it have not received much scholarly attention so far. Consequently, the Numeracy team faced several challenges during its work so far. A first challenge was to grapple with the definition of a slippery and expansive concept—numeracy—and to find ways to bound the range of tasks for assessing "numeracy" in a way that captures the breadth of the concept and yet is pragmatic for a large-scale assessment. Some areas, despite being part of our larger conception of numeracy (e.g., probabilistic reasoning) were not included in the item production process.

A related challenge stems from existence of overlap between scales. This is true regarding the Document Literacy scale, which includes some items requiring graph-reading or mathematical judgments, and the Problem-Solving scale, which does rely in some cases on tasks that call for mathematical operations. This situation reflects the fact that numerous everyday tasks present people with quantitative information or call for application of some mathematical skills in combination with other skills; mathematical information or skill demands cannot be artificially separated without hurting the ecological validity of assessment tasks. Our challenge was to reduce such overlap while maintaining coverage of key facets of numerate behavior.

A third challenge was to propose an approach for defining and assessing numeracy that not only agrees with "academic" conceptions of numeracy but that also will speak to the various stakeholders that should be interested in ILSS results—not only policymakers in different government bodies in participating countries, but also members of the vast communities involved in mathematics education and in literacy instruction, in schools, adult literacy or basic education programs, workplace training, and continuing education schemes. We believe that the tasks chosen, and the mathematical skills or knowledge they purport to capture, should be discussed in the same language used by teachers, trainers, and developers involved with mathematics or numeracy learning in these separate yet related contexts. Otherwise, it is possible that ILSS findings may have insufficient impact on educational day-to-day activities in classrooms and in all other venues where learning or improvement of numeracy is expected to take place.

On Assessing Mathematical Knowledge

Our goal here is to remind the reader of some differences between current assessment approaches, used either with adults or children, and explain the need to go beyond the Quantitative Literacy scale used in the IALS.

The IALS Quantitative Literacy Scale

To reiterate, the IALS framework made use of three literacy scales—Prose Literacy, Document Literacy, and Quantitative Literacy—to operationalize its conception of literacy.

Document Literacy (DL) was defined as:

> *The knowledge and skills required to locate and use information contained in various formats (including job applications, payroll forms, transportation schedules, maps, tables, and graphics).*

Quantitative Literacy (QL) was defined as:

> *The knowledge and skills required to apply arithmetic operations, either alone or sequentially, to numbers embedded in printed materials (such as balancing a check book, figuring out a tip, completing an order form, or determining the amount of interest on a loan).*

DL tasks required respondents to *identify*, *understand*, and *interpret* information given in various lists, tables, charts and displays; this information sometimes included quantitative information, such as numbers or percents. QL tasks required respondents to *apply arithmetical operations* (learned mostly in elementary grades). To be sure, QL tasks (as well as some DL tasks) did address important aspects of people's mathematical knowledge and skills. However, these tasks were limited in at least two important ways: (a) they related only to numerical information embedded in text, and hence did not require respondents to cope with other types of mathematical information (e.g., measurements, shapes) or with information whose processing does not require comprehension of rich text; and (b) they called for a very limited range of responses, i.e., exact computations or specific types of interpretations, and did not adequately represent the much wider range of types of responses that are typical of many everyday and work tasks, such as sorting, measuring, estimating, conjecturing, or using models (e.g., formulas).

In light of these limitations, it was necessary to depart from the IALS framework where Quantitative Literacy is viewed as a facet of literacy, and introduce Numeracy as a new and independent skill domain worthy of assessment as part of the broad conception of life skills adopted by ILSS. With this in mind, the following working definition of numeracy has been adopted:

> *The knowledge and skills required to effectively manage the mathematical demands of diverse situations.*

(This definition is presented here with little background, solely to provide a contrast with the definition of Quantitative Literacy above. Its full scope will not be apparent until much later, when we discuss details of the various facets of numeracy and of numerate behavior.)

School-Oriented Measures of Mathematical Knowledge

IALS' central mission was assessing facets of real-world literacy, hence the QL scale focused, as explained earlier, on application of basic mathematical operations in response to functional tasks using realistic *texts*. Yet, large-scale assessments of mathematical skills aimed at younger populations usually take quite different approaches. The brief review of some such assessments below intends to highlight some issues that can inform the content of an assessment scale of adult numeracy skills, but also shed light on areas where such an assessment has to deviate from familiar forms of assessments that are common with school-age populations.

The PISA survey currently being planned for OECD countries, for example, is developing a framework for assessing Mathematical Literacy, defined as follows:

> *An individual's ability to identify, to understand, to make well-founded judgements about, and to act towards the roles that mathematics plays in dealing with the world, as needed for that individual's current and future life as a constructive, concerned, and reflective citizen (PISA Mathematics Functional Expert Group, 1998).*

This definition shows some overlap and consistency with the conception of numeracy used in the present framework (as well as with broader conceptions of literacy as adopted by IALS and ILSS). Yet, some key differences seem to be the following:

- PISA focuses on how students understand, use, and apply mathematical skills and mathematize problems which are related to the *formal* school mathematics curriculum the students were expected to cover as part of their studies.
- PISA puts only partial emphasis on the realism of tasks (i.e., tasks can be contrived or use formal symbolism, given that students have limited world experience and are supposed to demonstrate mostly formal knowledge of what was taught to them).
- The PISA mathematical assessment is not explicitly interested either in tasks where mathematical information is embedded in text (realistic or otherwise), or in the influence of literacy skills on mathematical performance (despite the inclusion of the term "literacy" in "mathematical Literacy").

While different in several ways from the proposed PISA framework, the recent TIMSS study can also be characterized along the same lines in terms of its general emphasis.

Nationally-recognized standardized tests that are used in several countries to assess and qualify the mathematical knowledge of *adults* are often in line with school-based assessments. For example, the GED test in the U.S., which is used to grant a high-school equivalency diploma to adults who did not formally graduate from high-school, and the National Vocational Qualifications system in the U.K., both use items with characteristics that are in line with school-related assessments of mathematical knowledge than with the QL scale: they rely heavily on multiple-choice questions, employ tasks requiring manipulation of numbers without a meaningful context, or use some mathematical notations or assume some memorization of formulas.

The above discussion is not meant to be a comprehensive review of current large-scale assessments of schooling-related mathematical skills (see Robitaille & Travers, 1992). It simply reiterates that *all* assessments make conscious decisions regarding the (mathematical) skills that are important to assess, and consequently that the *forms* of assessment chosen carry not only advantages, but also disadvantages. The philosophy behind the design of mathematical assessments for TIMSS, GED, and similar assessments is based on assumptions about what it means to "know math" or "be able to do math" in a *schooling* context; hence, it is assumed that it is legitimate to use a certain degree of formalization of math symbols and to present contrived math tasks. This assumption does not fit the assessment of skills of adults who may have been out of school for many years or who have low literacy skills.

While school-oriented assessments are not meant for adults, they nevertheless point to some general areas of mathematical knowledge and skill that both school graduates as well as school leavers (or dropouts) may need to have to effectively cope with the various challenges of adult life. One lesson learned from reviewing the PISA, TIMSS, and similar assessments is that many important mathematical skills and knowledge were not captured by the QL scale. For instance, knowledge of "big ideas" related to shape and geometry or to chance and statistics, knowledge of measurement systems, or the ability to "model" the mathematical aspects of certain situations. Also, realistic tasks may be limited in terms of the kinds of mathematical concepts, ideas, or operations hidden in them—either higher-level mathematical skills (beyond the four basic operations, or simple percents) or mathematical knowledge and skills that are not dependent on textual information may exist. It follows that tasks involving only everyday situations and using only commonly-encountered texts (such as used in the QL) are somewhat limited in terms of the kinds of processes of mathematization that can be tapped through them.

Most curriculum frameworks for school mathematics education expect students to learn symbol-manipulation competencies without a clear connection to any real-life situation or task. Numeracy, however, is that bridge that links mathematics and the real-world. Our goal was to conceive of a range of tasks and stimulus material much broader in scope than that used in the QL scale, yet that do not rely on formal, school-based knowledge of mathematical systems. A challenge mentioned earlier in this regard was finding a balance between the use of text-based and relatively text-free tasks, while creating functional tasks with a reasonable degree of familiarity or realism. We wanted to do all this, while taking into account a view of "numeracy" based on assumptions about how *adults* "know" and "do math" in the real-world, based not only on their formal knowledge, to the extent it exists, but also on other experience-based knowledge that may have been acquired in informal contexts.

Adult Numeracy: Influences and Perspectives

This section briefly reviews several developments that have informed the conception of numeracy adopted for the purposes of creating the ILSS Numeracy scale.

Due to the interest in numeracy through workplace reform, the last few years have seen increased attention paid to numeracy of adults. Examples exist in many countries, such as the United Kingdom, the United States, Australia, the Netherlands, and various other OECD countries, and indicate that educational systems are recognizing the importance of numeracy. In the vocational area, basic skills projects in the United Kingdom and Australia developed frameworks that name numeracy as an important skill. These projects also identified properties that were associated with difficulty of numeracy tasks, given the need to describe stages or levels of accomplishment. For example, the National Council for Vocational Qualifications Core Skills in the United Kingdom identified five levels of numeracy skill (Oates, 1992) which closely linked to the sequence of content in the national school mathematics curriculum. Australia developed a set of generic "key competencies" for the vocational sector (Australian Education Council, 1992) , which covered three levels and included one numeracy/mathematics competency: "can use mathematical ideas and techniques"

In the adult education sector, which is growing and becoming more formalized in many countries due to economic considerations and the need for lifelong learning, attention to mathematical skills is very visible. Educators working with adults aim to assist students in developing mathematical concepts and relationships in ways that are personally meaningful but also functional. Adult educators usually assume that there is rarely one right way, but a wide variety of strategies that work well when solving functional computational problems. Students' idiosyncratic methods of using mathematics are encouraged and valued. This is often a significant difference from traditional (pre-reform) school-based mathematics teaching, within which school students are often expected to follow the one correct method, or algorithm, introduced by the teacher to solve a problem.

Realistic Mathematics Education

A major initiative in the Netherlands that started in the early 1980s was aimed mainly at school mathematics education (Heuvel-Panhuizen & Gravemeijer, 1991), but had influences also on adult education in numeracy. This perspective, called Realistic Mathematics Education (RME) is based upon the assumption that mathematics is an essential and important aspect of society, and therefore that mathematics education should be derived from real-life situations and should aim to create those skills applicable in any societal situation (family, work, etc). RME in adult education aims to optimize mathematical knowledge, skills, and problem-solving strategies that people have already been using in everyday life, or learned in or out of school, so that they can apply those strategies flexibly in all kinds of situations. RME also encourages adults to develop new knowledge, skills, and strategies that can enable them to have more control over their own personal, societal and work lives and undertake further learning or training. Therefore, in addition to mathematical knowledge and skills, adults are being trained in cooperative learning and in recognizing and facilitating their own problem-solving procedures, strategies, and learning processes (van Groenestijn, 1998).

The NCAL Framework

The Numeracy Project at the National Center for Adult Literacy (NCAL) in the U.S. has led to a certain perspective on adults' numeracy and on the facets of numerate behavior. It has been proposed (Gal, 1993) that "numeracy" refers to the aggregate of skills, knowledge, and dispositions that enable and support independent and effective management of diverse types of quantitative situations involving, for example, numbers, quantities, or visual or textual information that is based on mathematical ideas or has embedded mathematical elements.

Some situations call for *generative responses*, i.e., computing a number or generating an estimate or a decision. Examples are dealing with simple operations (measuring the length of a shelf), dealing with multi-step operations embedded in text (such as completing a tax form) and making reasonable decisions (for example, choosing the best loan). Other situations call for *interpretive responses*, i.e., making sense of quantitative statements or data displays (as in a newspaper article reporting crime statistics), and being able to ask critical questions about the information and arguments presented without performing any calculations. Both types of situations, and many mixed types, vary in terms of the literacy and communication skills they involve; in some cases it may not be possible to separate literacy from numeracy skills. It has also been suggested (Gal, 1997) that numerate behavior encompasses not only generative and interpretive skills but also *dispositional elements* (beliefs, attitudes, habits of mind) that motivate and support effective behavior in any given situation.

Gal (1993) argues that adults *manage* situations and do not "solve" them in ways that can be classified as right or wrong, in contrast to how students solve word problems, even if these are supposed to simulate real-life situations. For example, an adult who has to manage a generative situation (such as figuring the area of a room to carpet or a timetable for a trip) may ask herself/himself the following:

- How much time and energy do I have to spend on this?
- How accurate should I be?
- What is the cost of being slow or inaccurate?
- What tools (such as a calculator, a ruler, and/or paper and pen) can I use?

A person may decide to sacrifice precision or accuracy to reduce mental load or save time. A response may be reached in a computationally inefficient way or be based on non-standard procedures, but this may not matter in real-life as long as the individual manages the situation in a way that is reasonable in light of the demands of the situation and his or her goals. It follows that there may be important differences between demonstrating formal knowledge or understanding of mathematics on a school-oriented task (where adults may try to apply only school-based, memorized procedures), and demonstrating numerate, confident behavior in realistic situations. Cumming, Gal, & Ginsburg (1998) have argued that many of these aspects of numerate behavior have not been recognized by the mathematics education community and are not reflected in how tests and test items are created and interpreted.

The ANN Framework

The National Institute for Literacy in the United States has sponsored several efforts to define critical skill areas, as part of its *Equipped for the Future* initiative. One key project, by the Adult Numeracy Network (ANN) (Curry, Schmitt, and Waldron, 1996), was designed to reach a consensus on the kinds of mathematics that adults should know (and hence are important to teach in adult education). This project aimed to consolidate several perspectives, mainly those offered by the Curriculum Standards of the National Council of Teachers of Mathematics (NCTM) in the U.S., the Secretary's Commission on Achieving Necessary Skills (SCANS; U.S. Department of Labor, 1991), and prior work by the ABE Mathematics Team in Massachusetts (Leonelli, Merson, Schmitt, and Schwendeman, 1994), as well as the results of interviews with hundreds of adult learners, numeracy teachers, and employers.

The ANN's *Framework for Adult Numeracy Standards: The Mathematical Skills and Abilities Adults Need to be Equipped for the Future*, organized needed knowledge into seven broad themes/areas:

- Relevance/connections
- Problem solving/reasoning/decision making
- Communication
- Number and number sense
- Data
- Geometry: spatial sense and measurement
- Algebra: patterns and functions

The first three themes are concerned with processes of being numerate, while the latter four cover key content areas of mathematics. Again, this framework highlights and supports the view that numeracy is about making meaning of mathematical information.

Numeracy for Different Purposes

Since numeracy involves action in the real-world, it is important to reflect on the kinds of purposes served by people's ability to act in a numerate way.

There has been a range of work done in Australia to create a set of standards and a hierarchy of numeracy skill development that is not based upon school mathematics descriptions (Coates et al. 1995). In one such key project (Kindler et al. 1996), numeracy is organized into four broad categories or domains, according to different purposes and functions of using mathematics. *Numeracy for Practical Purposes* addresses aspects of the physical world that involve designing, making, and measuring. *Numeracy for Interpreting Society* relates to interpreting and reflecting on numerical and graphical information in public documents and texts. *Numeracy for Personal Organization* focuses on the numeracy requirements for personal organizational matters involving money, time and travel. *Numeracy for Knowledge* is another numeracy that describes the mathematical skills needed for further study in mathematics, or other subjects with mathematical underpinnings and/or assumptions.

These four purposes in general agree with the purposes identified for literacy. In Australia, for example, Kindler et al., (1996) reported on four such purposes: literacy for self expression, literacy for practical purposes, literacy for knowledge, and literacy for public debate. Work in the U.S. has described literacy in terms of four key related purposes: literacy as a vehicle for independent action; literacy as voice; literacy as bridge to the future; and literacy for access and orientation (Stein, 1995).

The separate aspects of the four broad purposes of numeracy identified by Kindler et al. (1996) are as follows:

- Numeracy for Practical Purposes—Measurement
- Numeracy for Practical Purposes—Design
- Numeracy for Interpreting Society—Numerical Information
- Numeracy for Interpreting Society—Data
- Numeracy for Personal Organization—Money
- Numeracy for Personal Organization—Time
- Numeracy for Personal Organization—Location
- Numeracy for Knowledge—Pattern and Algebra
- Numeracy for Knowledge—Problem-Solving

The IALS Quantitative Literacy and Document Literacy items appear to closely parallel Numeracy for Interpreting Society which focuses on dealing with numerical information and data embedded in texts. However, to validly represent the broader domain of numeracy, the ILSS numeracy scale needs to incorporate tasks that cover aspects of the other three domains of numeracy: Numeracy for Practical Purposes, Numeracy for Personal Organization and Numeracy for Knowledge. (Note that “practical purposes” and “personal organization” refer to functioning in work-related contexts *and* in home/shopping contexts—the actual mathematics underlying the demands in both types of contexts is quite similar.)

Numeracy: Definitions and Facets

We now turn to the broader issue of defining numeracy for the purposes of ILSS, and describing its facets or components. This is after having alerted above the reader to the need to (1) view numeracy as different from “knowing school mathematics” and from Quantitative Literacy as defined in IALS, and (2) appreciate the complexity of numerate behavior and of the multiple factors that affect the way adults cope with the demands imposed by tasks that contain mathematical or quantitative elements. (Note: For convenience, this report uses the term "mathematical" as inclusive of situations where *statistical* or *probabilistic* information may appear or where statistical thinking is required as well, even though statistical reasoning is not usually viewed a branch of mathematics).

Although there is neither a universally accepted definitions of numeracy nor agreement about how it differs from mathematics, most definitions contain an emphasis on the practical or functional application and use of mathematics. The Australian Beazley Committee definition is typical:

> *Numeracy is the mathematics for effective functioning in one's group and community, and the capacity to use these skills to further one's own development and of one's community (Beazley, 1984).*

Many authors argue that a discussion of functional skills should also address supporting or enabling attitudes and beliefs, an "at homeness" with numbers or "confidence" with mathematical skills, as these affect how skills and knowledge are actually put into practice.

> *Most important of all is the need to have sufficient confidence to make effective use of whatever mathematical skill and understanding is possessed (Cockcroft, 1982).*

Definitions of numeracy explicitly state that numeracy does not only refer to operating with numbers, as the word can suggest, but covers a wide range of mathematical skills and understandings. Buckingham (1997), for example, who studied what she called "specific and generic numeracies of the workplace" in some manufacturing industries in Australia, concluded that workplace numeracy is now about making decisions in the face of uncertainty in real situations, and that it encompasses far more than the basic skills traditionally associated with the term numeracy.

In recent years there has been much discussion and debate about the relationship between mathematics and numeracy and about the concept of "critical" numeracy. Johnston (1994), for example, has argued that:

> *To be numerate is more than being able to manipulate numbers, or even being able to 'succeed' in school or university mathematics. Numeracy is a critical awareness which builds bridges between mathematics and the real-world, with all its diversity (Johnston, 1994).*

Another important element in defining numeracy is that of the role of interpreting and communicating about the mathematics involved in different situations. Numeracy not only incorporates the individual's abilities to use and apply mathematical skills efficiently and critically, but also requires the person to be able to interpret and communicate about mathematical information and reasoning processes (Marr & Tout, 1997; Gal, 1997).

A definition that seems to incorporate much of the above aspects of numeracy is from the Queensland Department of Education (1994) in Australia:

> *Numeracy involves abilities that include interpreting, applying and communicating mathematical information in commonly encountered situations to enable full, critical and effective participation in a wide range of life roles.*

An important commonality in the above descriptions of numeracy is the presence of mathematical elements in real situations, and the notion that these can be used or addressed by a person in a goal-oriented way, dependent on the needs and interests of the individual within the given context (home, community, workplace, etc.), as well as on his or her dispositions. Our earlier discussions further imply that numeracy involves more than just procedural skills, but extend to possession of number sense, estimation skills, measurement, and to multiple ways of responding flexibly to a mathematical situation. Finally, given the extent to which numeracy pervades the modern world, it is not necessarily just *commonly* encountered situations that require numerate behavior, but also *new* situations.

With the above in mind, the definition of numeracy proposed for the ILSS is:

> ***Numeracy: The knowledge and skills required to effectively manage the mathematical demands of diverse situations.***

This definition of numeracy is much broader than the definition of QL (*The knowledge and skills required to apply arithmetic operations, either alone or sequentially, to numbers embedded in printed materials*); its key concepts relate in a broad way to situation management and to a range of effective responses to a range of mathematical demands (not only to application of procedural skills) to a wide range of skills and knowledge (not only to arithmetic operations) and to a wide range of situations that present actors with information with mathematical components (not only those involving *numbers* embedded in *printed* materials).

The next section elaborates on specific facets of numeracy that underlie the general terms used in the definition proposed for numeracy.

Facets of Numeracy

Building on ideas discussed earlier, we argue that numeracy-based behavior can be observed when a person has to satisfy a particular purpose in a real context and in this context has to manage the mathematical demands of a situation or task. People's numeracy is revealed through the responses or behaviors they generate (i.e., identifying, interpreting, acting upon, communicating) in reaction to the mathematical information or ideas that may be represented in a range of different ways and forms in different situations. The nature of a person’s responses to the mathematical demands of a situation will critically depend on the activation of various enabling knowledge bases (understanding of the context; knowledge and skills in the areas of mathematics, literacy, and statistics), on reasoning processes, and on certain dispositions. (Clearly, the expectation is that numerate behavior will involve an attempt to autonomously engage a task and not delegate it to others or deal with it by intentionally ignoring its mathematical content).

Thus, a proposed working definition of numerate behavior that will underlie the Numeracy Scale is:

> *Numerate behavior is observed when people* ***manage a situation or solve a problem in a real context****; it involves* ***responding*** *to* ***information about mathematical ideas*** *that may be* ***represented in a range of ways****; it requires the* ***activation of a range of enabling knowledge, behaviors, and processes.***

Table 1 lists specific components of the five key facets of numerate behavior (highlighted in the definition above). These facets and their components are further explained in subsequent sections.

Table 1: Numerate Behavior and Its Facets

Numerate behavior involves:

managing a situation or solving a problem in a real context
everyday life
work
societal
further learning

by responding
identifying or locating
acting upon
interpreting
communicating about

to information about mathematical ideas
quantity & number
dimension & shape
pattern and relationships
data & chance
change

that is represented in a range of ways
objects & pictures
numbers & symbols
formulae
diagrams & maps
graphs
tables
texts

and requires activation of a range of enabling knowledge, behaviors, and processes
mathematical knowledge and understanding
mathematical problem-solving skills
literacy skills
beliefs and attitudes.

Elaborations on Facets of Numeracy

A Purpose in a Real Context

People try to manage or respond to a numeracy situation because they want to satisfy a purpose or reach a goal. Four types of purposes and goals are described below. (To be sure, these are not mutually exclusive and may involve the same underlying mathematical themes).

Everyday life: The numeracy tasks that occur in everyday situations are often management tasks that one faces in personal and family life. Others revolve around hobbies, personal development, and interests.

Representative tasks are handling money and budgets, comparison shopping, personal time management, making decisions involving travel, planning holidays, mathematics involved in hobbies like quilting or wood-working, playing games of chance, understanding sports scoring and statistics, reading maps, and using measurements in home situations such as cooking or home repairs.

Work-related: At work, one is confronted with quantitative situations that often are more specialized than those seen in everyday life. In this context, people may develop good skills in managing situations that might be narrower in their application of mathematical themes. Representative tasks are completing purchase orders, totaling receipts, calculating change, managing schedules, using spreadsheets, organizing and packing different shaped goods, completing and interpreting control charts, making measurements, reading blueprints, tracking expenditures, predicting costs, and applying formulas.

Societal or community: Adults need to know about trends and processes happening in the world around them (e.g., regarding crime, wages, pollution, etc.) and may have to take part in social events or community action. This requires that adults can read and interpret quantitative information presented in the media, including statistical messages and graphs. Also, they may have to manage situations like organizing a fund raiser, realizing the fiscal effect of community programs, interpreting the results of a study of the latest health fad, and so forth.

Further learning: It is often also important to have numeracy skills that enable a person to participate in further study, whether for academic purposes or as part of vocational training. In either case, it is important to be able to understand symbols, rules and formulas, use a range of more formal mathematical skills, apply mathematical rules and principles, etc.

Responding

In different types of real-life situations, people may have to respond in one or more of the following ways (the first virtually always occurs; others will depend on the interaction between situational demands and the goals, skills, dispositions, and prior learning of the person):

1. **Identify or locate** some mathematical information present in the task or situation confronting them that is relevant to their purpose or goal
2. **Act** upon or react to the information in the situation. Bishop (1988), for example, proposed that there are six modes of mathematical actions that are common in all cultures: counting, locating, measuring, designing, playing and explaining. Other types of actions or reactions may occur, such as doing some calculations ("in the head" or with a calculator), ordering or sorting,

estimating, measuring, or modeling (such as by using a formula).

3. **Interpret** the information embedded within the situation (and the results of any prior action) and comprehend what it means or implies. This can include making a judgement about how mathematical information or known facts actually apply to the situation or context. This could mean that concetual judgment may have to be used, such as in deciding whether an answer makes sense or not in the given context, for example, that a result of "2.35 people" is not a real solution to how many people can actually fit into an average car.
4. **Communicate** about the mathematical information given, or the results of one's actions or interpretations to someone else. This can be done orally or in writing (a simple number or word, through to a more detailed explanation or analysis) and/or through drawing (a diagram, map, graph).

Information Related to Fundamental Mathematical Ideas

Mathematical information can be classified in a number of ways and on different levels of abstraction. One approach is to refer to fundamental "big ideas" in the mathematical world. Steen (1990), for example, identified six broad categories pertaining to: *Quantity, Dimension, Pattern, Shape, Uncertainty*, and *Change*. Rutherford & Ahlgren (1990) described networks of related ideas: *Numbers, Shapes, Uncertainty, Summarizing data, Sampling,* and *Reasoning*. Dossey (1997) categorized the mathematical behaviors of quantitative literacy as: *Data representation and interpretation, Number and operation sense, Measurement, Variables and relations, Geometric shapes and spatial visualization,* and *Chance*. We have drawn from these three closely tied categorizations to arrive at a set of five fundamental ideas that in our view characterize the mathematical demands met by adults in diverse situations at the beginning of the 21st century.

1. **Quantity and number**

Quantity is described by Fey (1990) as an outgrowth of people's need to quantify the world around us, using attributes such as: length, area, and volume of rivers or land masses; temperature, humidity, and pressure of our atmosphere; populations and growth rates of species; motions of tides; revenues or profits of companies, etc. **Number** is fundamental to quantification and different types of number constrain quantification in various ways: whole numbers can serve as counters or estimators; fractions, decimals and percents as expressions of greater precision, parts or comparisons (ratios); and positive and negative numbers as directional indicators. In addition to quantification, numbers are used to put things in order and as identifiers (e.g., telephone numbers or zip codes). Facility with quantity, number, and operation on number requires a good "sense" of magnitude. Contextual judgment comes into play when deciding how precise one should be or which tool (calculator, mental math, a computer) to use. Money and time management, the ubiquitous math that is part of every adult's life, depends on a good sense of number and quantity. A basic level numeracy task might be figuring out the cost of one can of soup, given the cost of 4 for $2.00; a task with a higher cognitive demand could require "harder numbers" when figuring out the cost per pound when buying 0.783 kg of cheese for 12,95 Euros.

2. **Dimension and shape**

Dimension includes “big ideas” related to one, two, and three dimensions of “things” (using spatial and numerical descriptions), projections, lengths, perimeters, planes, surfaces, location, etc. Facility with each dimension requires a sense of "benchmarks" and estimation, direct measurement and derived measurement skills. **Shape** is a category describing real images and entities that can be visualized (e.g., houses and buildings, packaging, snowflakes, knots, crystals, shadows and plants), as well as highly abstract “things” greater than three dimensions. Direction and location are fundamental qualities called upon when reading or sketching maps and diagrams. A basic numeracy task in this fundamental aspect could be shape identification whereas a complex task might involve describing the change in the size of an object when one dimension is changed.

3. **Pattern and relationships**

It is frequently written that mathematics is the study of patterns and relationships. **Pattern** is seen as a wide ranging concept that covers patterns encountered all around us, such as those in musical forms, ornamental design, natural geometry, traffic patterns, and objects we build. It is argued by Senechal (1990) that our ability to recognize, interpret, and create patterns is the key to dealing with the world around us. The human capacity for identifying **relationships** and for thinking analogically undergirds mathematical thinking. Algebra – beyond symbolic manipulation – provides a tool for representing relationships between amounts through the use of tables, graphs, symbols, and words. The ability to generalize and to characterize relationships between variables is a crucial gateway to understanding even the most basic economic, political or social analyses. A basic level numeracy task might require someone to describe how items are arranged in a package; developing a formula for an electronic spreadsheet would put a higher level of demand on the individual.

4. **Data and chance**

Data and chance encompass two related but separate topics. **Data** covers “big ideas” such as variability, sampling, error, or prediction, and related statistical topics such as data collection, data displays, and graphs. Modern society demands that adults interpret and produce organizers of data such as frequency tables, pie charts, graphs and sort out relevant from irrelevant data. **Chance** covers "big ideas" related to probability, subjective probability, and relevant statistical methods. Few things in the world are 100% certain; thus the ability to attach a number which represents the likelihood of an instance is a valuable tool whether it has to do with the weather, the stock-market, or the decision to board a plane. In this mathematical category, a simple numeracy skill might be the interpretation of a simple pie chart; a more complex task would be to infer the likelihood of an occurrence based upon past information such as predicting the weather.

5. **Change**

Change describes the mathematics of how the world changes around us. Individual organisms grow, populations vary, prices fluctuate, objects traveling speed up and slow down. Change and rates of change help provide a narration of the world as time marches on. Additive, multiplicative, exponential patterns of change can characterize steady trends; periodic changes suggest cycles and irregular change patterns connect with chaos theory. Describing weight loss compares as a simple task to calculating compounded interest.

Representation of Mathematical Information

Mathematical information in an activity or a situation may be available or represented in many forms. It may appear as concrete objects to be counted (e.g., people, buildings, cars, etc.) or as pictures of such things. It may be conveyed through mathematical symbols (e.g., numbers), notations, or formulas. Mathematical information may be encoded in a diagram or chart; similarly, a map of a real entity (e.g., of a city or a project plan) may contain information that can be quantified or mathematized. Likewise, graphs and tables may be used to display aggregate statistical or quantitative information (by displaying objects, counting data, etc.); further, a person may have to extract mathematical information from visual displays or from various types of texts, either in prose or in special formats (such as in tax forms).

Activation of Enabling Knowledge, Behaviors, and Processes

The way in which each person responds to the contexts, tasks, and mathematical representations discussed above will depend on the unique combination of existing mathematical knowledge, problem-solving skills and strategies, literacy skills, and dispositions that he or she brings to each situation. These are briefly discussed below.

Mathematical knowledge and understanding. The cognitive side of numerate behavior depends on an integration of mathematical knowledge, reasoning skills and problem-solving skills needed to be able to think and to act mathematically. Mathematical knowledge, including but not limited to the understanding of mathematical concepts and access to computational skills and procedures (formally learned or self-invented), is the basis for being able to manage many quantitative tasks in real-life situations. Such skills and procedures are detailed in many strands for school curricula and are assessed in many school-based surveys. A common set of such strands would include:

1. Whole numbers and the four basic operations
2. Ratios, percents, decimals, and fractions
3. Measurement
4. Geometry
5. Algebra
6. Probability and statistics

Problem-solving skills. Throughout life, adults develop or apply diverse strategies to manage their quantitative situations. Some strategies or skills may be based on prior formal learning, while others may be self-invented or adapted to fit the situation at hand. (See below about the importance of beliefs and prior experiences in this regard). To solve many computational problems or to figure a way to manage certain quantitative tasks, people have to re-construct reality in a mathematical way, for example, model or mathematize. They can do so either on their own or in discussion with other people. Problem-solving strategies may include, e.g., the ability to extract relevant information from the task/activity; rewriting/restating the task; drawing pictures, diagrams or sketches; guessing and checking; making a table; generating a concrete model or representation; etc.

Literacy skills. The ability to read, write, and talk are all important skills in undertaking a numeracy task or activity. Hence a person's literacy-related skills, reading strategies, and prior literacy experiences will impact on the person's numeracy performance and resulting numeracy level. In many cases understanding of "mathematical representations" will depend not only on formal mathematical or statistical knowledge but also on reading comprehension and other literacy skills. Numeracy tasks embedded in text may require a more analytical reading style than does prose. For example, following a computational procedure described in text (such as the text for computing shipping charges or adding taxes on an order form) may require special reading strategies, as text is very concise. Analyzing the mathematical relationships described in words requires specific skills, e.g., realizing that "four more than" is a different relationship than "four times as much."

Two different kinds of text may be encountered in functional tasks. The first involves the text that surrounds mathematical information of a numerical form—this is typical of all items from the IALS Quantitative Literacy scale. An example is a bank deposit slip with some text and instructions in which the numbers describing monetary amounts are embedded. The second involves cases where mathematical information is represented in textual form (i.e. words, phrases), but not with any mathematical notations. This may be as simple as conveying numerical data via number words, not digits (e.g., “five” instead of “5”). Seemingly basic mathematical or statistical terms may also appear (e.g., fraction, multiplication, percent, average, proportion). However, more complex textual representations of mathematical or statistical information can occur. Phrases such as “a function of,” “correlation,” “random fluctuations,” “representative sample,” “statistically significant,” or “rapid increase in crime rate,” and others can relate to mathematical ideas or phenomena without using familiar school-related terms, while mixing everyday and mathematical terms, and so forth. (Note: some Document Literacy items involve this second kind of representation of mathematical information). Obviously, both kinds of texts (i.e., surrounding text, and “mathematical” text) may be mixed within the same task, and may involve diverse mathematical concepts or ideas as well as various kinds of texts.

Beliefs, attitudes, and background experience. The way in which people respond to a quantitative situation and how they choose to act depends on how familiar they feel with such situations and how confident they are in their own strategies. General dispositions towards mathematical matters, as well as a person’s self-perception and the degree of a sense of “at-homeness” with numbers, considerably impact a person’s willingness and ability to perform mathematics tasks. It is well documented that math anxiety is quite common in some countries and this may be a crucial factor in the way that a person may respond to the mathematics embedded in a situation. Therefore, prior experiences and existing habits of coping with mathematical and numeracy situations may be influential.

In summary, it is not possible in our view to assess “numeracy” and determine the extent to which someone’s behavior is “numerate” by referring only to actual performance on test items. While this is a key goal of the ILSS, such assessment must be supplemented by the collection of other vital information about prior practices, attitudes, and other dispositional factors. Assessment of these variables through the background questionnaire is discussed later.

Complexity of Numeracy Items

Because of the scarcity of research on adults’ use and application of numeracy, there is little empirical knowledge that helps with defining what aspects make a numeracy activity or task more difficult or complex. The following section discusses this issue at length, given the centrality of difficulty/complexity factors to the analysis and interpretation of results from ILSS.

Previous Research on task complexity

In IALS, three factors were found to be the principal components of task difficulty (regarding literacy or text-based tasks): plausibility of distractors, type of match required, and type of information required. The difficulty of the Quantitative Literacy tasks appeared to be a function of several other factors:

1. The particular arithmetic operation required to complete the task
2. The number of operations needed to perform the task
3. The extent to which the numbers are embedded in printed materials
4. The extent to which an inference must be made to identify the type of operation to be performed (i.e. problem transparency; see below)

The IALS QL difficulty factors overall fit those used in large-scale assessments of mathematical skills (with children), which often make use of three or four factors:

1. *The mathematical concepts involved:* number systems and number sense, spatial and geometrical topics, functions and algebra, chance/statistics topics, etc. Concepts that are related to topics taught in earlier grades are considered easier.
2. *The complexity of operations*: addition, subtraction, multiplication, and so forth, as well as dealing with whole numbers, with decimals, with percents, etc. Operations that are related to topics taught in earlier grades are considered easier.
3. *The number of operations*: one-step problems are considered easier than multi-step problems.
4. *Problem transparency*: This factor is sometimes relevant; it refers to the extent to which the problem situation includes clearly identified numbers or entities and the extent to which it is clear what operations or actions to perform. (To the extent that these are not clear or transparent, respondents have to extract needed information by applying comprehension and inference strategies, making the task more complex.)

There are other adult-related assessment projects on which we can draw to develop the levels of complexity. Both the Essential Skills Research Project and the Applied Numeracy sub-test of the Work Keys test battery (American College Testing, 1997) use a two-factor model of complexity in their description of numeracy levels. The first factor is "operations required;" it is seemingly straightforward and refers to the difficulty of operations called for. However, this is complicated by the level of difficulty of the numbers being manipulated: computations that include fractions and decimals are usually more difficult than those with whole numbers. The Essential Skills model spells out two sequences of complexity on this factor: operations and translation.

A. Operations

1. Only the simplest operations are required and the operations to be used are clearly specified. Only one type of mathematical operation is used in the task.
2. Only relatively simple operations are required. The specific operations to be performed may not be clearly specified. Tasks involve one or two types of

mathematical operation. Few steps of calculations are required.

3. Task may require a combination of operations or multiple-applications of a single operation. Several steps of calculation are required. (More advanced operations may call for multiplication or division.)
4. Tasks involve multiple steps of calculation.
5. Tasks involve multiple steps of calculation. Advanced mathematical techniques may be required (e.g., percents, ratios, proportions).

The second complexity factor used in the Essential Skills model is the 'translation of information' (problem transparency) required, and is described as follows:

B. Translation (Problem Transparency)

1. Only minimal translation is required to turn the task into a mathematical operation. All the information required is provided.
2. Some translation may be required or the numbers needed for the solution may need to be collected from several sources. Simple formulae may be used.
3. Some translation is required but the problem is well defined.
4. Considerable translation is required.
5. Numbers needed for calculations may need to be derived or estimated; approximations may need to be created in cases of uncertainty and ambiguity. Complex formulae, equations or functions may be used.

Two considerations prompted us to question the appropriateness of using mathematics-related frameworks (from Essential Skills or elsewhere) as the sole source for development of a complexity scheme for items assessing *adults'* ability to cope with real-world numeracy tasks. First, effective coping with many real-world quantitative problems depends upon people's ability to make sense of and interact with different types of texts. This is hardly recognized by the Essential Skills model. Hence, it was essential to add difficulty factors that acknowledge the inherent links between literacy and numeracy, quite similar to those used in IALS.

Another, albeit more restricted consideration, is that the ordering of complexity of tasks by the type of operation performed may not be as clear with adults as it may be with children. Such ordering in school-based assessments is predicated on traditional school curricula, where more advanced topics are learned at higher grades. However, adults are known to use a lot of invented strategies, perhaps more so, and more efficiently so, than children. Multiplication or division problems, which can prove relatively hard for some young people, may be solved by (seemingly simpler) strategies, such as by repeated addition or repeated subtraction; complex numbers may be broken down in ways that ease mental load, and so forth. In addition, adults' familiarity with everyday contexts, such as with monetary entities, facilitates their performance with some seemingly advanced concepts.

For example, specific landmark values of fractions and percents, such as 1/2, 1/4, 50%, or 25%, are familiar to many people; as a result, they may be easier to manage than expected, violating curriculum-based ordering of difficulty. Hence, an *overall complexity level* has to be used, in order to weight these "inconsistencies" in ordering of difficulty levels proposed in other schemes.

With these considerations in mind, described below in more detail are five key factors that affect, separately and in interaction, the difficulty level of numeracy tasks to be used in the ILSS survey. A later subsection describes the calculation of an overall complexity level for each item, taking into account all five factors.

Complexity Factors: The ILSS case

The five factors of complexity outlined in Table 2 below are organized in two sets: three factors address the mathematical aspects of tasks, and two factors address mainly textual aspects of tasks. To be sure, these five factors are listed separately for clarity of presentation, but in actuality are *not* independent of each other and do interact in complex ways.

Table 2: Complexity Factors—Overview

Aspects	Category	Range
Mathematical aspects	1. Complexity of Mathematical information/data	Concrete/simple to abstract/complex
	2. Type of operation/skill	Simple to complex
	3. Expected number of operations	One to many
Textual aspects	4. Plausibility of distractors	No distractors to several distractors
	5. Type of match/problem transparency	Obvious/explicit to embedded/hidden

These five factors are discussed in the following sections. Descriptions of the three levels in each factor appear in Table 3 later on. (Notes: If a factor is irrelevant to an item, it is assigned a score of zero for that item. The description of some of the factors below may seem in cursory reading to repeat earlier discussions, but changes do appear, in part due to the broader emphasis here on non-computational and text-related items as well. We thus urge the reader to examine the full description of each complexity factor proposed for ILSS, to make sure our position at this point in the development phase is clear.)

Complexity of Mathematical Information: Some situations present a person with simple mathematical information, such as concrete objects (to be counted), simple whole numbers, or simple shapes or graphs. At lower skill levels, the information will be more familiar, whereas at higher levels, the information may be less familiar. Situations will be more difficult to manage if they involve more abstract or complex information, such as very large or very small numbers, decimals or percents, information about rates, or dense visual information, as in a diagram or complex table.

Type of Operation/Skill: Some situations require simple operations, such as addition or subtraction, or simple measurement (e.g., finding the length of a shelf). These are usually simpler than situations that require multiplication or division, and than situations that require manipulation of fractions or decimals. Percents may pose even greater difficulty. While the numerical system implied by a situation (be it additive, multiplicative, etc.) has direct bearing on task complexity, there may be exceptions, as noted above. Other tasks may combine both interpretive and generative skills and may involve other sequences of operations. The interpretation of information appearing in graphs, for example, may involve simple recognition of literal information, yet may become more complex if comparisons, conjecturing, or “reading beyond the information given” is required.

Expected Number of Operations: Tasks that require acting upon the mathematical information given may call for one application (step) of an operation, or for one action (e.g., literal reading of information in a table, or measurement). More complex tasks will demand multiple operations. Complex tasks may call for integration of information from several sources, or from numerical as well as text-based sources. Such information may be dispersed in different parts of a display, table, or text, and the person would have to perform multiple passes on the data or text to extract information or discover links between different pieces of data.

***Plausibility of Distract*ors:** This literacy-related variable has to do with whether or not given and/or requested information appears in the text. In general, tasks are easiest to process when there are no plausible distractors in the text. This usually happens when there is no other information in the text that meets any of the requirement of the task. At higher levels of difficulty, questions or tasks can involve irrelevant information both within the question as well as within the text. In terms of mathematical information, a low level of plausible distractors would mean that no other mathematical information was present apart from that requested, and numbers were easy to identify. At a higher level, there may be either some other mathematical information in the task (or its text) that could be a distractor, or mathematical information given or requested occurs in more than one place. A higher level of complexity could mean that outside information may well be needed to answer the question.

Type of Match/Problem Transparency: This is a combination of the factor of Problem Transparency outlined above, and of an IALS factor called Type of Match. Problem Transparency is a function of how well the mathematical information and tasks are specified and included aspects such as how the procedure is set out, how the values are made explicit or not, etc. Type of Match refers to the process that a respondent has to use to relate the requested action in the question to the information in the task or text, which can range from a simple action of locating or matching to more complex actions that require the respondent to perform a number of searches through the information given. This measure of complexity for a numeracy task incorporates the degree of text embededness of the mathematical information.

In easy tasks, the type of information (e.g., numerical values) and the operations needed are apparent and obvious from the way the situation is organized. In more difficult ones, the values must be located, derived, or calculated from other values; the operations needed may have to be discovered by the performer, depending on his or her interpretation of the context and of the kind of response expected. As well, numeracy situations may involve text to varying degrees, and this text may be of different degrees of importance. There may be situation where there is little or no text. Some situations may involve pure quantitative information that is to be interpreted or acted upon with virtually no text or linguistic input. In other words, the performer derives all the information needed to respond from the objects present in the situation or from direct numerical displays.

At a higher level, some textual or verbal information may be present alongside the mathematical information. The text can provide background information about the problem situation, or some instructions. For example, a bus schedule, cooking instructions, and a typical school-type word problem all involve some text and some numbers. Still other situations would be heavily text-based and may involve no numbers or mathematical symbols at all, just plain text. The task will contain mathematical or statistical information that a person needs to understand and, in some cases, act upon, but it will be much less transparent, heavily embedded in dense text or even requiring the need to use information from a number of sources within or even outside the text/task.

Overall Complexity Level

Each item will be scored on each of the five factors of complexity, according to the three levels described in Table 3 on the following page. An overall score will then be obtained for the complexity of an individual item by adding together the scores for each factor (yielding a total raw score in the range of 5 to 15).

The number of factors involved and the interplay among them will determine the overall difficulty or complexity level. Also, as tasks become more complex, actual performance on items may increasingly depend not only on the interplay among the five factors described here, but also on a person’s literacy skills and familiarity with formal mathematical concepts and terminology (both of which imply prior formal schooling). Performance on complex tasks will also depend on the person’s familiarity with the context in which a task is situated and on his or her background world knowledge, as well as on general problem-solving and reasoning skills. Thus, the difficulty of a task cannot in some cases be fully described without taking into account characteristics of the person who interacts with the task.

Table 3: Complexity Factors—Descriptions of Levels

Level 1 - score 1	Level 2 - score 2	Level 3 -score 3
Complexity of Mathematical Information/Data		
Information or answer required based on: - familiar real life activities, very concrete - Whole numbers in the range to 1,000s - benchmark fractions (1/2, 1/4) - decimal fraction or percentage for a half only (0.5 and 50%) - very simple whole number relations and patterns - standard monetary units - common everyday measures for length (whole units) - simple, common 2D shapes - simple graphs, tables, charts with few parameters and whole number values - time (dates, hours, minutes) - simple whole number data or statistical information - simple localized maps	Information or answer required based on: - common real life activities - large whole numbers in the range to millions - simple whole number rates and ratios - whole number relations and patterns - other benchmark fractions: 1/3, 1/4 1/10,.... - common decimals, like 0.1, 0.25 to 2 decimal places - common whole number percents, like 25% and 10%. - other everyday standard measures for length, weight and volume, including common fraction and decimal units - more complex 2D shapes,common 3D shapes - graphs, tables, charts with common data including whole number percents - more complex data or statistical information including whole number percents - common types of maps - whole number scales in 1s, 2s, 5s or 10s	Information or answer required based on: - all types of rational (and some irrational) numbers including directed numbers - formal mathematical information such as formulae - all kinds of measurement - complex shapes or combinations of shapes - complex graphs, tables or charts - complex data or statistical information
Complexity of Type of Operation/Skill		
a simple or straight forward activity or process that involves: - a simple arithmetical operation - locating/identifying data - making straight forward measurements - comparing or sorting values - following or giving straight forward directions	a more complex activity or process that involves: - more complex applications of the normal arithmetical operations such as calculating with rates, ratios, percentages, etc - estimating and rounding off - making and interpreting measurements - converting between standard measurement units - giving a simple explanation of a mathematical process	complex activity or process that involves: - more abstract and generative reasoning or explanations - more advanced mathematical techniques and skills e.g. indices and square roots - algebraic techniques - generating, organizing, graphing and interpreting grouped data - extrapolating data - converting between non-standard measurements
Complexity of Expected Number of Operations		
one operation, action or process	application of two or three steps, sometimes the same or similar operation, action or process	integration of several steps covering more than one different operation, action or process

Complexity of Plausibility of Distractors		
– no other mathematical information is present apart from that requested—no distractors - closed question—not open-ended	– there is some other mathematical information in the task that could be a distractor – the mathematical information given or requested can occur in more than one place – may need to bring to the problem simple information or knowledge from outside the problem. – fairly closed question	– more complex, open-ended task – mathematical information given or requested appears in several places. – other irrelevant mathematical information appears – necessary information may be missing, so outside information or knowledge may need to be brought in
Complexity of Type of Match/Problem Transparency		
The activity or operation required is clearly apparent and specified in little or no text, using objects and/or photographs or other clear, simple visualizations. Locating obvious information only	The activity can rely on text, using clear, simple sentences, and/or visualizations where some translation is required or where there is a need to locate information from a number of sources within the text/activity.	The activity is embedded in dense text where considerable translation is required or where values may need to be derived or estimated from a number of sources within or even outside the text/activity.

Item Development

Approach

Before describing the item pool that was developed, the following points should be noted:

1. Regarding facet 3, mathematical information: the actual content of the tasks was described in terms of common school-based mathematics topics and strands, i.e., whole numbers and basic operations; ratios, percents, decimals and fractions; measurement; geometry; algebra; and statistics. This is because the “big ideas” outlined above, while increasingly familiar to mathematics educators (and appearing in various publications as well as in the PISA Mathematics Framework), are more readily understood when mapped onto more familiar terms to educators.
2. With regard to the role of text in representations (facet 3), we noted above that there are two types of text involved: text conveying mathematical information (e.g., using number words or terms such as average), and surrounding text that provides information about the context of the problem and may at times function as a distractor. Item production aimed to generate not more than 1/3 of items that are embedded in rich text, and have about 2/3 of the items use little or no text. This was done to reduce overlap with the Document and Prose literacy scales, and to reflect the broad nature of numeracy tasks.
3. It is assumed that the other components of the enabling factors described earlier, i.e., problem-solving and dispositions, underlie all performance and are not factors that can be incorporated directly into the items. However, to reduce overlap with the Problem-Solving scale, tasks were kept short and are not

designed to simulate extended problem-solving processes.

4. Most tasks were derived from real-life stimuli or situations. Yet, to enable better assessment of respondents' ability to use mathematical models (e.g., a formula), some original stimuli were modified to support question-posing on such topics. Certain items were designed to overcome issues associated with the existence of different unit systems (e.g., monetary values, metric or Imperial measurements, etc), or to equate as much as possible the effect of the familiarity with specific unit systems in different countries, without changing the nature of the mathematical operations, whenever possible.
5. No context-free tasks were used, as quite often appear in school-based math surveys. Item content aims to be familiar to respondents across cultures. That said, the "realism," even of realistic tasks, may be limited and may not be fully known in advance. Relatively contrived tasks (e.g., not based on naturalistic stimuli) may be expected to be equally familiar to all respondents (as they have no culture-bound context) but in actuality may appear unfamiliar or "foreign" to some adults. Also, adults may be just as unfamiliar with "realistic" tasks taken from contexts other than those where they normally function.

Item Production

Development of Numeracy items used an item production grid with five key dimensions:

1. Type of purpose / context
2. Type of response
3. Type of mathematical information
4. Type of representation of mathematical information
5. Overall level of complexity (based on a combination of the five complexity factors discussed earlier)

Items were generated to cover key combinations of the first four dimensions (related to the facets described in Table 1, each with sub-components), with an emphasis on creating a sizable proportion of items that go well beyond the type of computational items used in the QL scale, thus representing the new and extended aspects of numeracy introduced in the present framework. The item production process also aimed to create tasks that spread across a range of difficulty levels, based on the complexity factors described. An effort was made to generate some items at the lowest level of complexity, which were only marginally covered by QL tasks in IALS, yet later received attention in some countries, as policies and educational programs may be earmarked for low-skill population.

In the first phase of development, over 100 items were developed. After internal review processes and some pilot testing, an item pool with 80 items was created and tested in the feasibility study (organized in 37 sets, where items in a set all relate to the same stimulus). Each item was assigned a complexity rating on a 10-point scale, using the five complexity factors described above; (this rating was later correlated with actual "p-values" (percent correct) to examine the validity of the complexity scheme).

The 80 items developed can be grouped or categorized in different ways, as they often involve more than one type of representation, information, response, or underlying mathematical reasoning or process. With that in mind, three crude yet informative categorizations of the items are as follows:

(Note: These categories are not mutually exclusive, so the total adds up to more than the total number of items: 68).

A. Item type:

Computational	32
Estimation	8
Measurement	8
Interpretive	28
Formula-related	4

B. Text in items:

Rich	9	sets
Some	11	sets
None/minimal	17	sets

C. Mathematical information involving:

Mathematical ideas:	***No. of items***
Quantity	63
Dimension/shape	29
Pattern/ Relationships	4
Data/chance	13
Change	14

While these counts are not the only way to categorize the nature of the item developed, they do suggest that a majority of the items created extend well beyond the type of items included in the QL scale. Further, a sizable proportion of the items classified as "computational" involve concepts or operations (e.g., regarding rates or specific types of percent problems) that go beyond what was covered by QL items, which focused mostly on application of the four basic operations. Further, many of the items have little if any text, and thus help in further separation of the terrain covered by the numeracy scale from that covered by the literacy scales.

Sample items that illustrate different types of tasks included in the initial item pool appear in Appendix 2.

Scoring

Coding guidelines were prepared for each item as part of the item development process. The basic four codes to be used in the field capture four conditions: answers that are "correct" or "wrong", as well as distinguish between "attempted/omitted" items (where any mark was found but not conclusive response) and "not attempted" items (were there is no indication the respondent tried to work on the problem).

For many items, multiple "correct" or "wrong" codes were initially prepared for use in the feasibility study, in order to capture specific types of errors or particular types of responses that are of interest in making sure that items elicit the expected type of responses (or mathematical activity/reasoning) and to make sure instructions are understood. In addition, in some items that require estimation or measurement, multiple codes were assigned to capture responses that may have different degree of accuracy yet still fall within the "correct" or "wrong" region, in order to collect more information about the level of accuracy or specificity that respondents adopt.

The scoring guidelines were compiled into a manual and used as part of the feasibility study of the numeracy scale, which was conducted in the U.S. and the Netherlands in June 1999 (see next section). Coding schemes which were initially created in English were translated to Dutch. Where necessary, different scoring rules were introduced in Dutch and English as part of the item development process to accommodate different unit systems that respondents may use (e.g., metric as opposed to Imperial measurements).

After the first groups of respondents were tested in the feasibility study, a sample of filled booklets was examined to collect realistic examples that can be listed in the coding/scoring manual, and to check the initial coding scheme against the actual range of real responses. Various modifications in scoring instructions were made in light of this stage.

Overall, the process of creating and adapting scoring instructions in two different languages, as well as the results of the feasibility study, have shown that most items created so far can be reliably translated, administered, and scored in more than one language while maintaining equivalency of meaning of item content and scoring instructions. Changes in some scoring instructions are now being examined in light of outcomes of the feasibility study, as described below. Also, ways of capturing if respondents use or do not use calculators are also explored.

A final point pertains to the possibility that specific codes capturing different types of responses that were considered either "correct" or 'wrong" will be retained for the field test and even for the main survey. The need to maintain specific codes may be helpful for several reasons.

The use of multiple codes or "rubrics" may enable the usage of partial credit, i.e., for a "partially correct" response. This feature is of high importance given that responses to certain numeracy tasks sometimes cannot be clearly categorized as correct or incorrect, but may fall along a range of levels of quality, or may show basic or partial understanding of a concept or procedure. Capturing information about partial knowledge states is valuable in light of the complex nature of numeracy, especially when non-computational (or interpretive) tasks are involved. In addition, retaining multiple response codes for selected items may render the data files much more useful for certain research endeavors, and could enable analysts and researchers to elicit much richer information about the nature of the numeracy skills that citizens possess (especially when secondary analyses are performed on the public data sets).

The use of multiple "correct" or "wrong" codes carries with it increased costs. At this point in time, the Numeracy Team is considering this issue in light of results from the feasibility study pertaining to the frequency of respondents who were assigned different response codes, and will present a coherent position in the near future for discussion within the project, before translation processes begin in all countries.

Capturing other "process" variables

The scoring rubrics designed for the Numeracy scale and described above are meant to capture basic aspects of the responses given, essential whether the response is correct or incorrect. Such information is the basis for determining numeracy levels in a nation or in specific subgroups. However, it is also important that countries are informed about reasoning processes and response strategies that people use when managing mathematical tasks, but that are not captured by kinds of scoring rubrics that can be pragmatically used in a large-scale survey. Examples are the strategies people employ when working on computational problems, quality of usage of calculators or tools, or the way people approach interpretive tasks, including those that are relevant in work contexts. Knowing more about such processes that are at the heart of numerate behavior can inform the planning of educational programs for improving numeracy levels. However, to date little or no large-scale surveys have explored these issues with adults, perhaps because a more qualitative analysis is needed.

While not part of the assessment planned by the ILSS Numeracy scale, the Numeracy Team sees it important to suggest to countries that it would be useful to plan small-scale qualitative studies in coordination with the main ILSS survey (perhaps using the same sampling process), in order to collect additional information on strategies, specific skills, and dispositions, especially of respondents at the lowest skill level ("level 1" in IALS). Such studies can be of much help in deciding on specific aspects of policies and programs that may emerge in response to the ILSS results on numeracy skills.

Background variables related to numeracy skills

A person may not necessarily act in numerate ways, even if she or he can demonstrate high ability on a numeracy test. The way in which a person responds to a numeracy task—including overt behavior or actions as well as cognitive processes and the propensity to adopt a critical stance—will depend not only on knowledge and skills but also on beliefs, attitudes, habits of mind, and prior practices.

Adult Numeracy Practices

Research suggests that, for adults as well as for children, mathematical knowledge develops both in and out of school (Lave, 1988; Saxe, 1991; Schliemann & Acioly, 1989; among others). Saxe and his colleagues have written about the importance of cultural practice in the development of mathematical thinking and how such practices profoundly influence an individual's cognitive constructions and mathematical ideas, depending, e.g., on the artifacts or tools they use, the nature of the measurement systems in their culture, the counting or calculating devices (abacus, calculator), the distribution of work among family members, or general patterns and types of social activity.

Mathematical experiences and practices, whether at work, home, when shopping, or in other contexts, can be both the result of a certain skill level, or the cause of skill levels. For this reason, we deemed it important to add to the ILSS Background Questionnaire (BQ) several items that examine the frequency of performing certain numeracy tasks in different contexts (in addition to those that were already used in the IALS BQ). Likewise, it was seen important to examine the frequency of use of artifacts such as calculators or computer spreadsheets and of tools (e.g., ruler) due to their possible relationship with numeracy skill.

Mathematical self-concept

In some cultures, some adults, including highly educated ones, decide that they are not "good with numbers." These sentiments or self-perceptions are usually attributed to negative prior experiences they have had as pupils of mathematics (Tobias, 1993), and stand in contrast to the desired sense of "at-homeness with numbers" (Cockcroft, 1982). Such attitudes and beliefs can interfere with one's motivation to develop new mathematical skills or to tackle math-related tasks, and may also affect test performance (McLeod, 1992). In realistic contexts, adults with a negative mathematical self-concept may elect to avoid a problem with quantitative elements, address only a portion of it, or prefer to delegate or subcontract a problem, e.g., by asking a family member or a salesperson for help. Such decisions or actions are indeed the prerogative of a manager and can serve to reduce both mental and emotional load. Yet, such actions may fall short of autonomous engagement with the mathematical demands of real-world tasks (as noted in the core definition of numeracy used here), carrying negative consequences, e.g., not being able to fully achieve one's goals, and thus

While the main goal of the Numeracy Team was to develop high-quality items, a separate, secondary goal was to identify items or scales assessing beliefs and practices related to mathematical competence. The logic behind developing these items was the expectation, supported by the professional literature as noted above, that actual numerate behavior may depend on the interaction between a person's skill level and his or her dispositions towards numeracy tasks or mathematical activities. people's tendency to enter (or avoid!) certain further learning opportunities (a topic of interest to policy makers and to the ILSS project) may be explained in part by their dispositions regarding learning or using mathematics, rather than by skill levels or demographic variables.

For these reasons, it appeared valuable to explore further new BQ scales as part of this feasibility study. Numerous items were developed and tested as part of the feasibility study. Preliminary data on this topic appear in the summary report from the feasibility study in Appendix 1. In light of these findings, tentative recommendations were recently passed to the ILSS Background Questionnaire Team, to include particular new belief- and practices-related items developed by the Numeracy Team and to retain certain related items from the IALS BQ.

Summary and next steps

Review of approach

The primary motivation for conducting the IALS was to inform policymakers of general levels of literacy skills and to understand factors associated with observed skill levels in participating countries. However, the IALS data pertaining to mathematical skills were limited to one specific area (Quantitative Literacy), that was never meant to assess a broad range of mathematical skills, and further made relatively heavy reliance on respondents' understanding of texts. The QL area, while very worthwhile as part of a broad conception of literacy, covers an important but still narrow part of Numeracy. Further, the QL scale also proved to be distant from the mainstream thinking of those involved in mathematics education at the school or adult levels. Hence, results from the IALS pertaining to Quantitative Literacy got little notice from those educational communities that can best act to close skill gaps involving mathematical knowledge.

The proposed numeracy scale will be based on a broad conception that acknowledges the information needs of policymakers but also considers concerns and accepted frameworks of the educational communities we expect to *react* to the outcomes of the survey. The perspective on the nature of numeracy presented in this framework document took into account the ideas that led to the development of the Quantitative Literacy Scale. Yet, the Numeracy Scale expands upon this early work in many ways.

The proposed Numeracy Scale includes not only computational tasks, but also tasks where quantitative information is not embedded in much text, or tasks where interpretations and opinions, as well as responses such as estimations or measurement rather than computation are required. While the new tasks do cover a much broader mathematical terrain compared to the QL, they do not assess the full range of numerate behavior, due to pragmatic considerations. Some aspects of people's numeracy skills cannot be fully reliably and validly assessed with the methodology presently available in the ILSS, especially those pertaining to problem-solving strategies or interpretive skills. On the other hand, the proposed numeracy tasks were designed with a stronger link to the way "knowing mathematics" or being "mathematically literate" is currently conceived by those involved in mathematics and numeracy education, and thus we do not foresee debate over their validity.

Numeracy is the bridge that links mathematical knowledge (whether acquired via formal or informal learning) with demands encountered in real-world situations that contain mathematical elements, and hence is a multifaceted and sometimes slippery construct. Numeracy demands of different situations are in part culture- and time-sensitive, thus possibly affecting how people from different countries approach or manage certain tasks. For example, differences in shopping habits (e.g., ubiquity of discounts in stores, usage of percentages in ads), the prevalence of hand-held calculators, the way different languages make it easy or difficult to handle or comprehend mathematical terms and words, or the cultural emphasis on finding accurate answers (as opposed to accepting estimates) can all influence the level of numeracy observed in different countries. The Numeracy Team designed tasks for the assessment scale with attention to features that can cause superficial differences between observed skill levels due to such cultural factors, but it is impossible to accommodate or eliminate all such differences without reducing the ecological validity of the tasks. It is expected that the process of adapting items in each country will help to improve some items in this regard.

Review of status of item pool

Overall, the feasibility study has shown that the item pool developed by the Numeracy Team includes items at diverse levels of difficulty that cover key facets of the conceptual framework for numeracy. Key factors that account for the difficulty of tasks were specified as part of the proposed conceptual framework. These factors, as reported in Appendix 1, proved to enable acceptable prediction of observed skill distributions; these factors thus can provide an important way for explaining the cognitive underpinnings of observed difference in numeracy performance of different people. While it was possible to complete only limited analyses so far due to time constraints, it overall appears that the results provide initial support both for the content validity and the construct validity of the numeracy scale. Further, relatively few serious problems were encountered during scoring processes in two languages, suggesting that, in general, the current item pool can be implemented in the field with few logistical and quality assurance problems.

That said, the feasibility study has pointed to some items that require revision (either the questions, the stimulus, or the scoring instructions). Further, several non-traditional items proved problematic and will have to be replaced. After deletion of items in light of the results from the IRT analysis, it appears that 6-10 new items will have to be developed, especially items requiring estimation, measurement, and contextual judgment, in order to achieve optimal coverage of those numeracy facets chosen for implementation in the ILSS.

However, in the recent ILSS project meeting at ETS (August 23-25, 1999) it was decided that in the field-test in 2001, *four* 30-minute blocks of numeracy items will be tested, each taking 30 minutes. This is an increase from the previous target of three blocks. (To be sure, in the main survey in 2002 Numeracy will still have only two 30-minute blocks). Assuming 20 items per block, this translates to an increase from 60 to 80 items in the item pool needed for the field-test. It was also deemed important to solicit additional input and suggestions for new items from participating countries so that the item pool to be field-tested in 2001 can reflect a broad consensus and representative types of items.

Next steps

In light of the above, over the next six months (By March 2000) the Numeracy Team will engage in three steps of work designed to provide the ILSS with a final item pool of 80 *tested* items, taking into account the need to work with other countries on contribution of new items, but at the same time keeping in mind the need to provide a sizable number of items early on so that countries can start adaptation and translation processes in a timely manner. Accordingly, the following tentative workplan is proposed:

First, by Dec. 1 1999, the Numeracy Team will:

1. Deliver 50 completed items (including revised scoring guidelines), drawn out of the current item pool that passed the feasibility study. These items will be accompanied by materials that describe critical elements and purpose of each item in order to support proper translation and adaptation processes.

2. Develop 6-10 new items in areas deemed important (see above)

3. In parallel, work with participating countries to facilitate the contribution of new items (with their scoring guidelines). The team will provide training and support to those experts who will be chosen by participating countries to take part in this process, in order to focus this development stage on those types and levels of items that are most needed in order to round-up the coverage of key numeracy facets. Items and scoring guidelines proposed by participating countries will be reviewed by the Team.

4. Conduct a peer-review of both the conceptual framework and some existing items by colleagues from several countries who are familiar with adult education or workplace training needs. (Note: This review process was planned by the Team on request from the ILSS management *before* details of the process of contribution of additional items by participating countries were known. It is felt that reliance on both sources, i.e., officially assigned experts from participating countries, and other peers recognized within the numeracy and adult education or training communities (some from non-participating countries), will be beneficial and can further contribute to the credibility and value of the Numeracy scale).

Second, in January-February 2000, the Numeracy Team will conduct a limited test trial of the new items proposed so far on samples of 30-50 people in the US and NL. The items to be tested will be selected from those developed by the Team and by experts from contributing countries, after all new items were reviewed for fit with the current framework and needed difficulty levels.

Finally, by March 1, the Numeracy Team will deliver 30 additional items needed to complete the item pool (with scoring guidelines, translation support materials, and administration protocol). These items will be selected from a pool comprised of the new items that passed the limited test trial, and the remaining original items that passed the feasibility study but were not selected among the first 50 delivered by Dec 1.

The above three-stage process can satisfy the need to provide high-quality items early on to participating countries, yet leaves enough room for adoption of needed new items, while expecting all new items to be tested under realistic and equal conditions and satisfy the same quality and psychometric criteria.

Authors

Iddo Gal	*University of Haifa, Israel*
Dave Tout	*Language Australia, Australia*
Mieke van Groenestijn	*Hogeschool van Utrecht, the Netherlands*
Mary Jane Schmitt	*National Center for the Study of Adult Learning and Literacy, Harvard University, U.S.*
Myrna Manly	*El Camino College, California, U.S.*

References

American College Testing. (1997). *Work keys: Preliminary technical handbook*. Iowa City, IA: Author.

Australian Education Council (The Mayer Committee). (1992). *The key competencies report: Putting general education to work.* Canberra: The Australian Education Council and Ministers for Vocational Education, Employment, and Training.

Beazley, K. (1984). Education in Western Australia: Report of the Committee of Inquiry into Education in Western Australia. Education Department of Western Australia.

Bishop, A. (1988). Mathematical enculturation: a cultural perspective in mathematics education. Dordrecht, Holland: D. Reidel Publishing Co.

Buckingham, E. (1997). *Specific and generic numeracies of the workplace*. Melbourne: Deakin University.

Coates et al. (1995). *National Reporting System*. Australia: Commonwealth of Australia and the Australian National Training Authority.

Cockcroft, W.H. (1982). Report of the Committee of Inquiry into the Teaching of Mathematics in Schools. London: HMSO.

Cumming, J., Gal, I., and Ginsburg, L. (1998). *Assessing mathematical knowledge of adult learners: Are we looking at what counts?* (Technical Report TR98-05). Philadelphia, PA: University of Pennsylvania, National Center on Adult Literacy.

Curry, D., Schmitt, M.J., and Waldron, S. (1996, July). A framework for adult numeracy standards: The mathematical skills and abilities adults need to be equipped for the future. Boston, MA: The Adult Numeracy Practitioners Network.

Dossey, J.A. (1997). "Defining and measuring quantitative literacy". In L.A. Steen (Ed.), *Why numbers count: Quantitative literacy for tomorrow's America*. New York: College Entrance Examination Board.

European Commission. (1996). *White paper on education and training: Teaching and learning towards the learning society*. Luxembourg: Office for Official Publications of the European Commission.

Fey, James T. (1990). "Quantity" In L.A. Steen (Ed.) *On the shoulders of giants: New approaches to* *numeracy*. Washington, DC: National Academy Press.

Gal, I. (1993). *Issues and challenges in adult numeracy* (Technical Report TR93-15). Philadelphia, PA: National Center on Adult Literacy, University of Pennsylvania.

Gal, I. (1997). Numeracy: Imperatives of a forgotten goal. In L. A. Steen (Ed.), *Why numbers count: quantitative literacy for tomorrow's America* (pp. 36-44). New York: The College Board.

van Groenestijn, M. (1998). Constructive numeracy teaching as a gateway to independent learning. In Coben, D., and O'Donoghue, J. (Eds.), *Adults learning mathematics-4, Proceedings of the Fourth International Conference of Adults Learning Mathematics*. London: Goldsmiths University of London.

Heuvel-Panhuizen, M. van den, and Gravemeijer, K.P.E. (1991). Tests are not all that bad: An attempt to change the appearance of written tests in mathematics instruction at the primary school level. In L. Streefland (Ed.), *Realistic mathematics education in primary school*. The Netherlands: Utrecht.

Johnston, B. (1994, Summer). Critical numeracy? In *Fine print*, Vol.16, No. 4.

Jones, S. (1995). The distribution of literacy. In *Literacy, economy and society*. Ottawa: Statistics Canada.

Kindler, Jan, et al. (1996). *Certificates in General Education for Adults*. Victoria, Melbourne: Adult, Community and Further Education Board.

Lave, J. (1998). Cognition in practice: Mind, mathematics, and culture in everyday life. Cambridge, UK: Cambridge University Press.

Leonelli, E., Merson, M.W., Schmitt, M.J., and Schwendeman, R. (Eds.). (1994). *The Massachusetts ABE math standards project* (Vols. 1 and 2). Holyoke, MA: Holyoke Community College.

Marr, B., and Tout, D. (1997). A numeracy curriculum: Australian Association of Mathematics Teachers (AAMT) conference proceedings. Melbourne: AAMT.

Mathematics Functional Expert Group-PISA Study, Jan de Lange, Chair. (1998, June). *Framework for assessing mathematical literacy*. Draft technical report submitted to the Organisation for Economic Cooperation and Development (OECD)/Program for International Student Assessment (PISA) project.

McLeod, D.B. (1992). Research on affect in mathematics education: A reconceptualization. In Grouws, D.A. (Ed.), Handbook of research on mathematics teaching and learning (pp. 575-596). New York: Macmillan.

Murnane, R., Willett, J., and Levy, F. (1995, March). *The growing importance of cognitive skills in wage determination* (Working Paper No. 50-76). Cambridge, MA: National Bureau of Economic Research.

National Council of Teachers of Mathematics (NCTM). (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

Oates, T. (1992). Developing and piloting the NCVQ core skill units: An outline of method and a summary of findings (Report 16). London: National Council on Vocational Qualifications (NCVQ).

Robitaille, D. F., and Travers, K. J. (1992). International studies of achievement in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 687-709). New York: Macmillan.

Rutherford, F.J and Ahlgren, A. (1990). *Science for all Americans*. New York: Oxford University Press.

Saxe, G.B. (1992). Culture and cognitive development: Studies in mathematical understanding. Hillsdale, NJ: Lawrence Erlbaum Associates.

Saxe, G.B., Dawson, V., Fall, R., and Howard, S. (1996). Culture and children's mathematical thinking. In Sternberg, R, and Ben-Zeev, T. (Eds.), *The nature of mathematical thinking* (pp. 119-144). Hillsdale, NJ: Lawrence Erlbaum Associates.

Schliemann, A.D., and Acioly, N.M. (1989). Mathematical knowledge developed at work: The contribution of practice versus the contribution of schooling. *Cognition and Instruction*, 6, 185-221.

Secretary's Commission on Achieving Necessary Skills (SCANS). (1991). *What work requires of schools: A SCANS Report for America 2000*. Washington, D.C.: U.S. Dept. of Labor.

Senechal, Majorie (1990) "Shape" In L.A. Steen (Ed.) *On the shoulders of giants: New approaches to numeracy*. Washington, DC: National Academy Press.

Statistics Canada, and The Organisation for Economic Cooperation and Development (OECD). (1996). *Literacy, economy, and society: First results from the International Adult Literacy Survey.* Ottawa: Author.

Statistics Canada, Human Resources Development Canada, and The Organisation for Economic Cooperation and Development (OECD). (1997). *Literacy skills for the knowledge society: Further results from the International Adult Literacy Survey*. Ottawa: Author.

Steen, L. (Ed.). (1990). *On the shoulders of giants: New approaches to numeracy*. Washington, D.C.: National Research Council.

Stein, S.G. (1995). Equipped for the future: A customer-driven vision for adult literacy and lifelong learning. Washington, D.C.: National Institute for Literacy.

Tobias, S. (1993). *Overcoming math anxiety*. New York: Norton.

Willis, S. (1990). *Being numerate: What counts?* Victoria: Australian Council for Educational Research (ACER).

Appendix C.1 - Numeracy: Feasibility Study Report

This study was planned with 3 goals in mind:

- *Examine items*: the performance of individual items, error patterns, clarity of instructions and stimulus material, and clarity of scoring schemes, in two different languages.
- *Examine the scale*: Collect evidence that will show how well the numeracy scale is capable of measuring the construct it is designed to measure, and data about the distribution of item difficulty levels (p-values) within different subgroups.
- *Identify informative background variables*: Explore new questions developed by the Numeracy Team that can contribute to the explanation of performance on the Numeracy scale and to the prediction of trends of interest to ILSS (e.g., participation in further learning activities), beyond what is offered by other variables already planned for the ILSS Background Questionnaire (BQ).

Design

Samples: The study was conducted in the United States (US) and the Netherlands (NL), and was planned to include samples of at least N=300 in each country, so as to provide a minimum of 100 responses for each item tested. Samples were to be balanced by age, education level, and gender. Recruitment was conducted by Westat in the US and NIPO in the NL, and used contacts with diverse employers and education providers, as well as using pre-existing phone listings of marketing research and survey firms. Testing sessions took place in Washington D.C. and Amsterdam over a period of 3 weeks in June 1999 and respondents were primarily from these urban areas.

Procedures: Due to cost considerations, all materials were administered in groups, using paper and pencil booklets to which subjects responded in writing. A modest incentive was offered to participants. Sessions were conducted in groups of 10-40. Most respondents stayed for about 1.5 hours, but in 5 sessions in each country 10-12 respondents stayed for an additional 30-minute focus group in which comments on certain items and suggestions for changing items or were solicited.

Materials: Each respondent received two booklets after an initial session orientation:

Background Booklet: This booklet was allocated 15 minutes and involved two types of items: Biographical data and prior schooling (17 items), and experimental attitudes and practices scales (43 items).

Test booklet: This booklet included 32-33 test items, and was allocated 60 minutes. Respondents also received a can containing a ruler, 4-function calculator, and simulated Euro money, and could use these items as needed. Respondents also recorded starting and ending times and answered a brief follow-up written survey regarding their comfort level, use of calculator, and comments about problematic items.

Construction of test booklets: A total of 84 items were tested in the feasibility study. These 84 items were arranged in 5 Blocks, each with 15-17 items, equated by expected level of complexity. Test booklets were then created by using 2 blocks per booklet, so that each block appeared in two booklets, once as the first block and one as the second block, to control for order effects. The 84 items included 80 new items developed by the Numeracy Team, and 4 “anchor items” taken from the QL scale of the IALS and planned to enable calibration of the results from the current samples in comparison to results for the same items in nationally representative US and NL samples used in IALS.

Scoring, coding, analysis: Scoring was performed in ways intended to approximate conditions in the main study. Scorers received a training workshop and a scoring manual with explanations of codes and sample responses for all items (derived from examination of filled booklets in US and NL). All booklets were scored independently by two scorers. Disagreements were recorded, and resolved through arbitration by a third scorer. Data were captured manually into computer files. Raw data files were analyzed by members of the Numeracy Team, and recoded files (using correct-incorrect codes) were also sent to ETS for IRT analysis by Dr. Kentaro Yamamoto, who previously analyzed the IALS data.

Results

Note: Data files from the feasibility study were received on the last week of August. As only preliminary analyses were completed until this document went to print, some of the following results may have to be interpreted with caution, yet overall are useful for evaluation of the characteristics of the Numeracy items and scale.

Sample characteristics. A total of 754 respondents were sampled. Table 1 contains a general description of the samples. As can be seen, the US sample, in comparison to the NL sample, included an oversampling of subjects with higher levels of education and undersampling of young subjects and of males. Otherwise, samples appeared quite comparable and represent all gender, education, and age groups. (More subjects were sampled in the middle age group due to its wider range)

Table 1: Characteristics of samples

Variable		**US**	**NL**
		N=386	N=368
Gender	Male	39%	49%
	Female	61%	51%
Education	Below HS	17%	31%
	HS	39%	31%
	Above HS	44%	38%
Age	16-24	26%	29%
	25-54	50%	42%
	55-65	24%	29%

Testing time. Each booklet was answered by about 70-75 persons in each country. Each item appeared in two booklets and thus was answered by about 150 persons. The average time per booklet ranged from 42-48 minutes for 32-33 items. Assuming an average response rate of 1.5 minutes per item, it means that in the main study, in which each numeracy block will be given 30 minutes testing time, it may be possible to employ blocks with approximately 20 items, or a total of up to 40 numeracy items.

IRT analyses. IRT analyses were performed on all items and results were presented during the last Team meeting in Los Angeles on August 17-21. For each item, several indices were computed, mainly: (a) point-biserial correlations between performance on item and on total booklet score (indicating, informally, the extent to which the item measures what is being measured by the rest of the scale); (b) p-value (percent correct); and (c) several other IRT parameters (e.g., difference scores, deviations from expected item curve, differential performance in the two samples). Based on these indices, a total of 34 items were flagged as possibly problematic, while 46 demonstrated acceptable values on all indices. (The 4 anchor IALS items were excluded from analyses). Of the 34 flagged items, it was decided to drop 12 items, fix 10 items, and keep 12 “as is” (given that differential performance in the two samples was explained by translation errors or other known differences between the countries (e.g., ease of using metric system). Thus, a total of 68 items appear to be “good” (“as is” or after known revision) at this stage of the analysis. (Note: this number may change once additional sources of data are examined).

Distribution of p-values. After deletion of problem items, the distribution of p-values (percent correct) was derived for the 68 “good” items, as shown in table 2.

Table 2: Distribution of p-values (68 “good” items)

p-value	# items
0% - 39%	9
40%-49%	11
50%-59%	4
60%-69%	12
70%-79%	16
80%-89%	11
90%+	5

To aid in interpreting this distribution, it is noted that the current sample performed on the 4 IALS anchor items about 10% higher (on average) compared to performance levels in the US and NL for the same items in IALS. Informally, this means that the p-values observed in the feasibility study may be overestimating by roughly 10%, on average, compared to what can be expected if the samples were nationally representatives. The higher performance level in the present study was expected as both samples did not include a reasonable proportion of subjects with low levels of education or literacy. The data in Table 2 overall suggest that the current item pool, after initial deletion of problem items, is capable of covering a wide range of p-values, as it includes items that range from easy to relatively difficult.

Correlation of p-values with complexity ratings. As explained earlier, the numeracy team has developed a preliminary conceptual framework that includes five factors which “drive” item complexity. Based on these 5 factors, each with three levels, a total “difficulty score” in the range 5-15 was pre-assigned by the team to each item before the data from the feasibility study were known. A key statistic to emerge from the feasibility study was the correlation between this total “complexity rating” and the actual p-value of all items. This correlation was computed for the 68 “good” items and was found to be r=0.77. This high correlation indicates that the conceptual system of complexity factors proposed by the Numeracy Team can actually account for at least 50% of the performance on the numeracy items. This finding implies that the scores to be obtained on the numeracy scale can be interpreted quite well by using this 5-factor system, and thus that the distribution of ability levels on the numeracy scale can be meaningfully explained by reference to these factors. That said, the numeracy team is now exploring items whose actual p-value was quite different from the difficulty level predicted by the complexity factors, both as a way to identify problem items (or instructions) as well as a way to refine the understanding of the factors that make numeracy tasks vary in difficulty.

Patterns of performance. This section reports additional evidence regarding the potential of the present item pool to identify differences between national samples or population subgroups that may be of value to policy makers. Such patterns are important as they can show if the total booklet score (which is the best estimate available for the ability level of each person, using classical test theory conception) appears to “behave” as can be expected based on prior known patterns related to national, age, or gender differences. However, due to time constrains prior to submission of this early report, computations of total booklet scores was conducted before IRT results (see above) were presented by the ETS specialist to the Numeracy Team. Total score for each respondent was computed as the percent correct on all items in a booklet, including items that were later deleted. Hence, results below should be interpreted with caution and will be redone in the near future.

Table 3: Total score (percent correct) in two countries, total and by subgroups

Variable		**US**	**NL**
		N=386	N=368
Total		57%	67%
Gender	Male	63%	71%
	Female	54%	65%
Education	Below HS	37%	57%
	HS	56%	68%
	Above HS	66%	75%
Age	16-24	63%	77%
	25-54	55%	69%
	55-65	56%	58%

Some patterns of interest in Table 3 are:

- The NL sample performs about 10% higher than the US sample. This statistic is roughly comparable to differences between NL and US samples in IALS, and also agrees with data from TIMMS showing advantage to Dutch students over US students.
- Men perform somewhat better than women. This finding agrees with various reported results showing advantage to males over females in school populations, and with some IALS results.
- Performance improves with level of education. This finding agrees with normative expectations for impact of education on mathematical knowledge. Yet, even people with low education level appear able to complete quite a few numeracy tasks. While the feasibility sample is biased in this area (i.e., doesn't have enough people with very low education levels and low literacy levels), the scale does show potential for judging the impact of education on skill, but also shows that the correlation between the two is far from being high.
- Performance declines with age, but not in the same rate, across the two samples. Again, this finding is in part in line with previous findings from IALS.

To be sure, the two samples should not be compared to each other as the samples' characteristics are somewhat different, and the definition of education levels was not identical due to differences in education systems in US and NL. Overall, the patterns of performance observed in Table 3 agree with previously reported findings, though also show some new trends. While these results will have to be re-generated after item deletion, they nonetheless do appear to support other data reported earlier and tentatively suggest that the numeracy scale is measuring what it is designed to measure.

New BQ scales. A separate goal of the feasibility study was to evaluate the predictive value of new scales developed by the numeracy team, comprised of items assessing beliefs and practices related to mathematical topics. The logic behind developing these items was the expectation, supported by the literature, that actual numerate behavior may depend on the interaction between a person's skill level and his or her dispositions towards numeracy tasks or mathematical activities. Such dispositions could result from various beliefs about one's skill level, or depend on current mathematical practices at home or at work. Further, people's tendency to enter (or avoid!) certain further learning opportunities (a topic of interest to policy makers and to the ILSS project) may be explained in part by their dispositions regarding learning or using mathematics, rather than by skill levels or demographic variables. For these reasons, it appeared valuable to explore further new BQ scales as part of this feasibility study.

Preliminary findings so far suggest that there would be merit in including in the ILSS BQ a factor composed of three items centering around an adult's confidence and comfort with everyday mathematical demands. This factor has high internal consistency (alpha over 0.80) and has a high correlation (r=.51) with numeracy skill level (i.e., percent of correct responses). Some predictive power for this factor remains even when controlling for education level, age, and gender which are known to be predictive of literacy level. An adult who scores high on this factor (F3Q20) strongly agrees that s/he is good with numbers and calculations [20j], has the math skills needed to do his/her current job well [20l], and does not feel anxious when figuring discounts when shopping [20m]. The original IALS BQ included related self report items: E8 (How would you rate your mathematical skills for your main job?) and E9 (To what extent are your mathematical skills limiting your job opportunities- for example, advancement or getting another job?). These two items were found in IALS to correlate with the QL score.

Other scales being explored deal, for example, with how comfortable a person is using technology (a calculator, a computer, and spreadsheets), or examine the frequency of certain mathematical practices at work, such as filling out tables and graphs, using a calculator, measuring, and more. While analyses have not been conduced yet, preliminary recommendations have been made to the BQ development team to adopt certain items that correlate highly with the overall numeracy score.

Summary

Overall, the feasibility study has shown that the item pool developed by the Numeracy Team includes items at diverse levels of difficulty that cover key facets of the conceptual framework for numeracy. Performance on these items varies in subgroups in ways that agree with prior literature on mathematical skills, and is explainable by theoretical factors specified as part of the conceptual framework. Hence, while only preliminary analyses have been completed so far of the collected data, it overall appears that the results provide initial support both for the content validity and the construct validity of the scale. Further, relatively few serious problems were encountered during scoring and coding processes, suggesting that, in general, the current item pool can be implemented in the field with relatively few logistical quality assurance problems. Finally, the feasibility study has shown that the item types and formats shown for administration are short enough so that it will be possible to fit up to about 40 items in the main study in the time allotted for Numeracy blocks. This number of items will ensure a reasonable coverage of all critical facets of numeracy.

In coming weeks, the Numeracy Team will complete additional statistical analyses as well as selectively examine additional qualitative and quantitative data about items, such as (a) statistics about scorer disagreement and comments by scorers; (b) focus-group comments on items; (c) comments from the proctors of the group testing sessions; and (d) comments by respondents on specific items they had trouble with. Based on these analyses, items or scoring guidelines in need of revision will be identified and modified as needed.

Finally, the feasibility study has shown the need to develop 6-10 new items, with a focus on the areas of Measurement and Estimation. These areas proved to be harder to assess by the items included in the feasibility study, in part because they appeared to respondents to be unusual for math-related assessments, and left room for responses with many different degrees of accuracy depending on respondents perception of the “demand characteristics” of the tasks at hand. Hence, more scoring problems were encountered on such items as well. Development of such new items will ensure adequate coverage of critical facets of numeracy in the main survey.