Annex A

A Construct-Centered approach to Understanding What was Measured in the Adult Literacy and Life Skills (ALL) Survey

Table of Contents

Annex A

A Construct-Centered approach to Understanding What was Measured in the Adult Literacy and Life Skills (ALL) Survey

Overview

This annex offers a brief overview of the frameworks that were used to develop and interpret the scales used to measure prose and document literacy, numeracy, and problem solving in the Adult Literacy and Life Skills (ALL) survey. The importance of developing a framework is thought to be central in construct-based approaches to measurement. Among the things that should be included in any such framework are an agreed upon definition of what ought to be measured and the identification of characteristics that can be used in the construction and interpretation of tasks. In addition to describing these characteristics for each measure, this annex also includes sample items along with the identification of item features that are shown to contribute to item difficulty. Collectively this information provides a means for moving away from interpreting survey results in terms of discrete tasks or a single number and towards identifying levels of performance sufficiently generalized to have validity across assessments and groups.

Introduction

In 1992, the Organization for Economic Co-operation and Development (OECD) (OECD, 1992) concluded that low literacy levels were a serious threat to economic performance and social cohesion on an international level. But a broader understanding of literacy problems across industrialized nations – and consequent lessons for policy makers – was hindered due to a lack of comparable international data. Statistics Canada and Educational Testing Service (ETS) teamed up to build and deliver an international comparative study of literacy.

The International Adult Literacy Survey (IALS) was the first comparative survey of adults designed to profile and explore comparative literacy distributions among participating countries. In 2000, a final report was released (OECD and Statistics Canada, 2000) which included the results from three rounds of

assessments involving some 23 country/language groups representing just over 50 per cent of the world's GDP. While IALS laid an important foundation for international comparative surveys of adults, there were also calls to expand what was being measured. There was a growing concern among governments and policy makers as to what additional competencies are relevant for an individual to participate fully and successfully in a modern society and for a society to meet the challenges of a rapidly changing world. One project aimed at addressing this issue was entitled *Definition and Selection of Key Competencies* (DeSeCo) and was carried out under the leadership of Switzerland. Its goal was to lay out, from a theoretical perspective, a set of key competencies that are believed to contribute to a successful life and a well-functioning society (Rychen and Salganik, 2003).

In response to these calls for broader measures, the ALL survey commissioned the development of frameworks to use as the basis for introducing new measures into the comparative assessments of adults. Those responsible for the development of ALL recognized that the design of any reliable and valid instrument should begin with a strong theoretical underpinning that is represented by a framework that characterizes current thinking in the field. According to Messick (1994) any framework that takes a construct-centered approach to assessment design should: begin with a general definition or statement of purpose – one that guides the rationale for the survey and what should be measured in terms of knowledge, skills or other attributes; identify various performances or behaviours that will reveal those constructs, and; identify task characteristics and indicate how these characteristics will be used in constructing the tasks that will elicit those behaviours.

This annex provides an overview of the frameworks used to develop tasks that measure prose and document literacy, numeracy, and problem solving in the ALL survey. In characterizing these frameworks this annex also provides a scheme for understanding the meaning of what has been measured in ALL and for interpreting levels along each of the scales. It borrows liberally from more detailed chapters that were developed in conjunction with the ALL survey (Murray, Clermont and Binkley, in press).

Scaling the literacy, numeracy and problem solving tasks in ALL

The results of the ALL survey are reported along four scales – two literacy scales (prose and document), a single numeracy scale, and a scale capturing problem solving – with each ranging from 0 to 500 points. One might imagine these tasks arranged along their respective scale in terms of their difficulty for adults and the level of proficiency needed to respond correctly to each task. The procedure used in ALL to model these continua of difficulty and ability is Item Response Theory (IRT). IRT is a mathematical model used for estimating the probability that a particular person will respond correctly to a given task from a specified pool of tasks (Murray, Kirsch and Jenkins, 1998).

The scale value assigned to each item results from how representative samples of adults in participating countries perform on each item and is based on the theory that someone at a given point on the scale is equally proficient in all tasks at that point on the scale. For the ALL survey, as for the IALS, proficiency was determined to mean that someone at a particular point on the proficiency scale would have an 80 per cent chance of answering items at that point correctly.

Just as adults within each participating country in ALL are sampled from the population of adults living in households, each task that was constructed and used in the assessment represents a type of task sampled from the domain or construct defined here. Hence, it is representative of a particular type of literacy, numeracy or problem solving task that is associated with adult contexts.

One obvious question that arises once one looks at the distributions of tasks along each of the described scales is, what distinguishes tasks at the lower end of each scale from those in the middle and upper ranges of the scale? Do tasks, that fall around the same place on each scale share some set of characteristics that result in their having similar levels of difficulty? Even a cursory review of the items reveals that tasks at the lower end of each scale differ from those at the higher end.

In an attempt to display this progression of complexity and difficulty, each proficiency scale was divided into levels. Both the literacy and numeracy scales used five levels where Level 1 represents the lowest level of proficiency and Level 5 the highest. These levels are defined as follows: Level 1 (0-225), Level 2 (226-275), Level 3 (276-325), Level 4 (326-375) and Level 5 (376-500). The scale for problem solving used four levels where Level 1 is the lowest level of proficiency and Level 4 the highest. These four levels are defined as follows: Level 1 (0-250), Level 2 (251–300), Level 3 (301–350), and Level 4 (351–500).

Since each level represents a progression of knowledge and skills, individuals within a particular level not only demonstrate the knowledge and skills associated with that level but the proficiencies associated with the lower levels as well. In practical terms, this means that individuals performing at 250 (the middle of Level 2 on one of the literacy or numeracy scales) are expected to be able to perform the average Level 1 and Level 2 task with a high degree of proficiency. A comparable point on the problem solving scale would be 275. In ALL, as in IALS, a high degree of proficiency is defined in terms of a response probability of 80 (RP80).[1] This means that individuals estimated to have a particular scale score are expected to perform tasks at that point on the scale correctly with an 80 per cent probability. It also means they will have a greater than 80 per cent chance of performing tasks that are lower on the scale. It does not mean, however, that individuals with given proficiencies can never succeed at tasks with higher difficulty values; they may do so some of the time. It does suggest that their probability of success is "relatively" low – i.e., the more difficult the task relative to their proficiency, the lower the likelihood of a correct response.

An analogy might help clarify this point. The relationship between task difficulty and individual proficiency is much like the high jump event in track and field, in which an athlete tries to jump over a bar that is placed at increasing heights. Each high jumper has a height at which he or she is proficient – that is, the jumper can clear the bar at that height with a high probability of success, and can clear the bar at lower heights almost every time. When the bar is higher than the athlete's level of proficiency, however, it is expected that the athlete will be unable to clear the bar consistently.

Measuring prose and document literacy in ALL

Defining prose and document literacy

The National Adult Literacy Survey (NALS), which was funded by the National Center for Education Statistics (NCES) as part of its overall assessment program in adult literacy, was the largest and most comprehensive study of adult literacy ever conducted in the United States (Kirsch, Jungeblut, Jenkins and Kolstad, 1993). Like all large-scale assessments funded by NCES, NALS was guided by a committee, which was comprised of a group of nationally recognized scholars, practitioners, and administrators who adopted the following definition of literacy:

> *"Literacy is using printed and written information to function in society, to achieve one's goals, and to develop one's knowledge and potential."*

This definition captures the initial work of the committee guiding the development of the assessment and provides the basis for creating other aspects of the framework to be discussed. It was also reviewed and adopted by the countries participating in the first round of IALS and was carried forward in ALL. This definition includes several assumptions made by panel members and, thus, it is important to consider various parts of this definition in turn.

Beginning with "*Literacy is…*", the term literacy is used in preference to "reading" because it is likely to convey more precisely to a non-expert audience what the survey is measuring. "Reading" is often understood as simply decoding, or reading aloud, whereas the intention of the adult surveys is to measure something broader and deeper. Researchers studying literacy within particular contexts noted that different cultures and groups may value different kinds of literacy practices (Sticht, 1975; Heath, 1980; Szwed, 1981). Heath, for example, found that uses for reading could be described in terms of instrumental, social interactional, news-related, memory supportive, substitutes for oral messages, provision of a permanent record, and personal confirmation. The fact that people read different materials for different purposes implies a range of proficiencies that may not be well captured by signing one's name, completing a certain number of years of schooling, or scoring at an 8th-grade level on a test of academic reading comprehension.

The phrase "… *using printed and written information*" draws attention to the fact that panel members view literacy not as a set of isolated skills associated with reading and writing, but more importantly as the application of those skills for specific purposes in specific contexts. When literacy is studied within varying contexts, diversity becomes its hallmark. First, people engage in literacy behaviours for a variety of uses or purposes (Sticht, 1978; Heath, 1980; Cook-Gumperz and Gumperz, 1981; Mikulecky, 1982). These uses vary across contexts (Heath, 1980; Venezky, 1983) and among people within the same context (Kirsch and Guthrie, 1984a). This variation in use leads to an interaction with a broad range of materials that have qualitatively different linguistic forms (Diehl, 1980; Jacob, 1982; Miller, 1982). In some cases, these different types of literacy tasks have been associated with different cognitive strategies or reading behaviours (Sticht, 1978, 1982; Crandall, 1981; Scribner and Cole, 1981; Kirsch and Guthrie, 1984b).

The phrase "… *to function in society, to achieve one's goals, and to develop one's knowledge and potential*" is meant to capture the full scope of situations in which literacy plays a role in the lives of adults, from private to public, from school to work, to lifelong learning and active citizenship. "To achieve one's goals and to develop one's knowledge and potential" points to the view that literacy enables the fulfillment of individual aspirations—those that are defined such as graduation or obtaining a job, and those less defined and less immediate which extend and enrich one's personal life. The phrase "to function in society" is meant to acknowledge that literacy provides individuals with a means of contributing to as well as benefiting from society. Literacy skills are generally recognized as important for nations to maintain or improve their standard of living and to compete in an increasingly global market place. Yet, they are equally as important for individual participation in technologically advancing societies with their formal institutions, complex legal systems, and large government programs.

Identifying task characteristics

The task characteristics represent variables that can be used in a variety of ways in developing an assessment and interpreting the results. Almond and Mislevy (1998) have identified five roles that variables can take on. They can be used to limit the scope of the assessment, characterize the features that should be used for constructing tasks, control the assembly of tasks into booklets or test forms, characterise examinees' performance on or responses to tasks, or help to characterise aspects of competencies or proficiencies. IALS focused on variables that can be used to help in the construction of tasks as well as in the characterization of performance along one or more proficiency scales.

Each task in the assessment represents a piece of evidence about a person's literacy (Mislevy, 2000). While the goal of the assessment will be to develop the best possible picture of an individual's skills and abilities, the test cannot include an infinite number of tasks nor can an infinite number of features of those tasks be manipulated. Therefore, decisions need to be made about which features should be part of the test development process. Three task characteristics were identified and used in the construction of tasks for the IALS. These characteristics include:

Adult contexts/content. Since adults do not read written or printed materials in a vacuum, but read within a particular context or for a particular purpose, materials for the literacy assessment are selected that represent a variety of contexts and contents. This is to help ensure that no one group of adults is either advantaged or disadvantaged due to the context or content included in the assessment. Six adult context/content categories have been identified as follows:

- Home and family: may include materials dealing with interpersonal relationships, personal finance, housing, and insurance.
- Health and safety: may include materials dealing with drugs and alcohol, disease prevention and treatment, safety and accident prevention, first aid, emergencies, and staying healthy.
- Community and citizenship: may include materials dealing with staying informed and community resources.
- Consumer economics: may include materials dealing with credit and banking, savings, advertising, making purchases, and maintaining personal possessions.

- Work: may include materials that deal in general with various occupations but not job specific texts, finding employment, finance, and being on the job.
- Leisure and recreation: may include materials involving travel, recreational activities, and restaurants.

Materials/texts. While no one would doubt that a literacy assessment should include a range of material, what is critical to the design and interpretation of the scores that are produced are the range and specific features of the text material which are included in constructing the tasks. A key distinction among texts that is at the heart of the IALS survey is their classification into continuous and non-continuous texts. Conventionally, continuous texts are formed of sentences organized into paragraphs. In these texts, organization occurs by paragraph setting, indentation, and the breakdown of text into a hierarchy signalled by headings that help the reader to recognize the organization of the text. The primary classification of continuous texts is by rhetorical purpose or text type. For IALS, these included: expository, descriptive, argumentative, and injunctive.

Non-continuous texts are organized differently than continuous texts and so allow the reader to employ different strategies for entering and extracting information from them. On the surface, these texts appear to have many different organizational patterns or formats, ranging from tables and schedules to charts and graphs, and from maps to forms. However, the organizational pattern for these types of texts, which Mosenthal and Kirsch (1998) refer to as documents, is said to have one of four basic structures: a simple list; a combined list; an intersected list; and a nested list. Together, these four types of documents make up what they have called matrix documents, or non-continuous texts with clearly defined rows and columns. They are also closely related to other non-continuous texts that these authors refer to as graphic, locative, and entry documents.

The distinction between continuous and non-continuous texts formed the basis for two of the three literacy scales used in IALS. Continuous texts were the basis for tasks that were placed along the prose scale while non-continuous texts formed the basis for tasks along the document scale. The quantitative scale included texts that were both continuous and non-continuous. The distinguishing characteristic for this scale was that respondents needed to identify and perform one or more arithmetic operations based on information contained in the texts. This scale was replaced in ALL with the numeracy scale, which is discussed in more detail later in this annex.

Processes/strategies. This task characteristic refers to the way in which examinees process text to respond correctly to a question or directive. It includes the processes used to relate information in the question (the given information) to the necessary information in the text (the new information) as well as the processes needed to either identify or construct the correct response from the information available. Three variables used to investigate tasks from national and international surveys will be summarized here. These are: type of match, type of information requested, and plausibility of distracting information.

Type of match

Four types of matching strategies were identified: locating, cycling, integrating, and generating. *Locating* tasks require examinees to match one or more features of information stated in the question to either identical or synonymous information

provided in the text. *Cycling* tasks also require examinees to match one or more features of information, but unlike locating tasks, they require respondents to engage in a series of feature matches to satisfy conditions stated in the question.

Integrating tasks require examinees to pull together two or more pieces of information from the text according to some type of specified relation. For example, this relation might call for examinees to identify similarities (i.e., make a comparison), differences (i.e., contrast), degree (i.e., smaller or larger), or cause-and-effect relations. This information may be located within a single paragraph or it may appear in different paragraphs or sections of the text. In integrating information, examinees draw upon information categories provided in a question to locate the corresponding information in the text. They then relate the text information associated with these different categories based upon the relation term specified in the question. In some cases, however, examinees must *generate* these categories and/or relations before integrating the information stated in the text.

In addition to requiring examinees to apply one of these four strategies, the type of match between a question and the text is influenced by several other processing conditions which contribute to a task's overall difficulty. The first of these is the number of phrases that must be used in the search. Task difficulty increases with the amount of information in the question for which the examinee must search in the text. For instance, questions that consist of only one independent clause tend to be easier, on average, than those that contain several independent or dependent clauses. Difficulty also increases with the number of responses that examinees are asked to provide. Questions that request a single answer are easier than those that require three or more answers. Further, questions which specify the number of responses tend to be easier than those that do not. For example, a question which states, "List the 3 reasons…" would be easier than one which said, "List the reasons…". Tasks are also influenced by the degree to which examinees have to make inferences to match the given information in a question to corresponding information in the text, and to identify the requested information.

Type of information requested

This refers to the kinds of information that readers need to identify to answer a test question successfully. The more concrete the requested information, the easier the task is judged to be. In previous research based on large-scale assessments of adults' and children's literacy (Kirsch and Mosenthal, 1994; Kirsch, Jungeblut, and Mosenthal, 1998), the type of information variable was scored on a 5-point scale. A score of one represented information that was the most concrete and therefore the easiest to process, while a score of five represented information that was the most abstract and therefore the most difficult to process.

For instance, questions which asked examinees to identify a person, animal, or thing (i.e., imaginable nouns) were said to request highly concrete information and were assigned a value of one. Questions asking respondents to identify goals, conditions, or purposes were said to request more abstract types of information. Such tasks were judged to be more difficult and received a value of three. Questions that required examinees to identify an "equivalent" were judged to be the most abstract and were assigned a value of five. In such cases, the equivalent tended to be an unfamiliar term or phrase for which respondents had to infer a definition or interpretation from the text.

Plausibility of distractors

This concerns the extent to which information in the text shares one or more features with the information requested in the question but does not fully satisfy what has been requested. Tasks are judged to be easiest when no distractor information is present in the text. They tend to become more difficult as the number of distractors increases, as the distractors share more features with the correct response, and as the distractors appear in closer proximity to the correct response. For instance, tasks tend to be judged more difficult when one or more distractors meet some but not all of the conditions specified in the question and appear in a paragraph or section of text other than the one containing the correct answer. Tasks are judged to be most difficult when two or more distractors share most of the features with the correct response and appear in the same paragraph or node of information as the correct response.

Characterizing prose literacy tasks

There are 55 tasks ordered along the 500-point prose literacy scale representing 19 IALS prose literacy tasks and 36 new prose literacy tasks designed and developed for the ALL survey. These tasks range in difficulty value from 169 to 439. One of the easiest tasks (receiving a difficulty value of 188 and falling in Level 1) directs the reader to look at a medicine label to determine the "maximum number of days you should take this medicine." In terms of our process variables, type of match was scored as easy because the reader was required to locate a single piece of information that was literally stated in the medicine label. The label contained only one reference to number of days and this information was located under the label dosage. Type of information was scored as easy because it asked for a number of days and plausibility of distractor was judged to be easy because there is no other reference to days in the medicine label.

MEDCO ASPIRIN 500

INDICATIONS: Headaches, muscle pains, rheumatic pains, toothaches, earaches. RELIEVES COMMON COLD SYMPTOMS.

DOSAGE: ORAL. 1 or 2 tablets every 6 hours, preferably accompanied by food, for not longer than 7 days. Store in a cool, dry place.

CAUTION: Do not use for gastritis or peptic ulcer. Do not use if taking anticoagulant drugs. Do not use for serious liver illness or bronchial asthma. If taken in large doses and for an extended period, may cause harm to kidneys. Before using this medication for chicken pox or influenza in children, consult with a doctor about Reyes Syndrome, a rare but serious illness. During lactation and pregnancy, consult with a doctor before using this product, especially in the last trimester of pregnancy. If symptoms persist, or in case of an accidental overdose, consult a doctor. Keep out of reach of children.

INGREDIENTS: Each tablet contains
500 mg acetylsalicicylic acid.
Excipient c.b.p. 1 tablet.
Reg. No. 88246

Made in Canada by STERLING PRODUCTS, INC.
1600 Industrial Blvd., Montreal, Quebec H9J 3P1

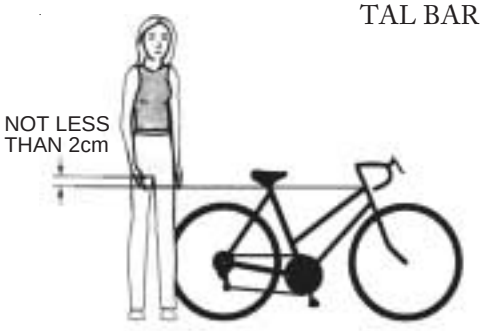
0 67736 11079

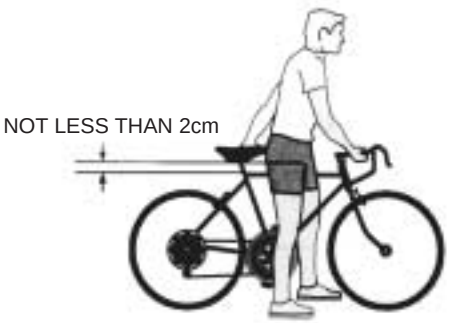
Reprinted by permission

A second prose literacy task directs the reader to look at an article about impatiens. This task falls in the middle of Level 2 and has a difficulty value of 254. It asks the reader to identify "what the smooth leaf surfaces and the stems suggest about the plant." Again, the task directed the reader to locate information contained in the text so it was scored easy for type of information. The last sentence in the second paragraph under the heading *Appearance* states: "The smooth leaf surfaces and the stems indicate a great need of water." Type of information was scored as being moderate because it directs the reader to identify a condition. Plausibility of distractor was scored as being moderate also because the same paragraph contained a sentence which serves to distract a number of readers. This sentence states, "… stems are branched and very juicy, which means, because of the tropical origin, that the plant is sensitive to cold."

PROPER FRAME FIT

RIDER MUST BE ABLE TO STRADDLE BICYCLE WITH AT LEAST 2 cm CLEARANCE ABOVE THE HORIZONTAL BAR WHEN STANDING.





NOTE: Measurement for a female should be determined using a men's model as a basis.

PROPER SIZE OF BICYCLE

FRAME SIZE	LEG LENGTH OF RIDER
430mm	660mm-760mm
460mm	690mm-790mm
480mm	710mm-790mm
530mm	760mm-840mm
560mm	790mm-860mm
580mm	810mm-890mm
635mm	860mm-940mm

OWNER'S RESPONSIBILITY

1. **Bicycle Selection and Purchase:** Make sure this bicycle fits the intended rider. Bicycles come in a variety of sizes. Personal adjustment of seat and handlebars is necessary to assure maximum safety and comfort. Bicycles come with a wide variety of equipment and accessories . . . make sure the rider can operate them.
2. **Assembly:** Carefully follow all assembly instructions. Make sure that all nuts, bolts and screws are securely tightened.
3. **Fitting the Bicycle:** To ride safely and comfortably, the bicycle must fit the rider. Check the seat position, adjusting it up or down so that with the sole of rider's foot on the pedal in its lowest position the rider's knee is slightly bent.

Note: Specific charts illustrated at left detail the proper method of deter-mining the correct frame size.

The manufacturer is not responsible for failure, injury, or damage caused by improper completion of assembly or improper maintenance after shipment.

Tasks which fall at higher levels along the scale present the reader with more varied demands in terms of the type of match that is required and in terms of the number and nature of distractors that are present in the text. One such task (with a difficulty value of 281 or the beginning of Level 3) refers the reader to a page from a bicycle's owner's manual to determine how to ensure the seat is in the proper position. Type of information was scored as moderate because the reader needed to identify and state two conditions that needed to be met in writing. In addition, they were not told how many features they needed to provide from among those stated. Type of information was also scored as moderate also because it involved identifying a condition and plausibility of distractor received a score indicating it was relatively easy.

A somewhat more difficult task (318), one near the top of Level 3, involves an article about cotton diapers and directs the reader to "list three reasons why the author prefers to use disposable rather than cotton diapers." This task is made more difficult because of several of our process variables. First, type of match was scored as difficult because the reader had to provide multiple responses, each of which required a text-based inference. Nowhere in the text does the author say, "I prefer cotton diapers because…". These inferences are made somewhat more difficult because the type of information being requested is a "reason" rather than something more concrete. This variable also was coded as difficult because of its abstractness. Finally, plausibility of distractor was scored as moderate because the text contains information that may serve to distract the reader.

An additional task falling in Level 4 on the Prose literacy scale (338) directs the reader to use the information from a pamphlet about hiring interviews to "write in your own words one difference between the panel and the group interview." Here the difficulty does not come from locating information in the text. Rather than merely locating a fact about each type of interview, the reader needs to integrate what they have read to infer a characteristic on which the two types of interviews differ. Experience from other surveys of this kind reveal that tasks in which readers are asked to contrast information are more difficult, on average, than tasks in which they are asked to find similarities. Thus, type of match was scored as complex and difficult. Type of information was scored as being difficult as well because it directs the reader to provide a difference. Differences tend to be more abstract in that they ask for the identification of distinctive or contrastive features related in this case to an interview process. Plausibility of distractor was judged as being easy because no distracting information was present in the text. Thus this variable was not seen as contributing to the overall difficulty of this task.

The Hiring Interview

Preinterview

Try to learn more about the business. What products does it manufacture or services does it provide? What methods or procedures does it use? This information can be found in trade directories, chamber of commerce or industrial directories, or at your local employment office.

Find out more about the position. Would you replace someone or is the position newly created? In which departments or shops would you work? Collective agreements describing various standardized positions and duties are available at most local employment offices. You can also contact the appropriate trade union.

The Interview

Ask questions about the position and the business. Answer clearly and accurately all questions put to you. Bring along a note pad as well as your work and training documents.

The Most Common Types of Interview

One-on-one: Self explanatory.

Panel: A number of people ask you questions and then compare notes on your application.

Group: After hearing a presentation with other applicants on the position and duties, you take part in a group discussion.

Postinterview

Note the key points discussed. Compare questions that caused you difficulty with those that allowed you to highlight your strong points. Such a review will help you prepare for future interviews. If you wish, you can talk about it with the placement officer or career counsellor at your local employment office.

The most difficult task on the prose literacy scale (377) falls in the lower range of Level 5 and required readers to look at an announcement from a personnel department and to "list two ways in which CIEM (an employee support initiative within a company) helps people who lose their jobs because of departmental reorganization." Type of match was scored difficult because the question contained multiple phrases that the reader needed to keep in mind when reading the text. In addition, readers had to provide multiple responses and make low text-based inferences. Type of information received a moderate score because readers were looking for a purpose or function and plausibility of distractor was scored as relatively difficult. This task is made somewhat more difficult because the

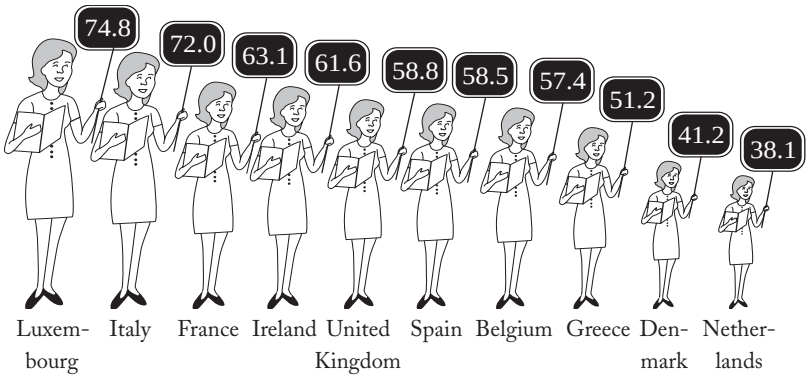
announcement is organized around information that is different from what is being requested in the question. Thus while the correct information is listed under a single heading, this information is embedded under a list of headings describing CIEM's activities for employees looking for other work. Thus, this list of headings in the text serves as an excellent set of distractors for the reader who does not search for or locate the phrase in the question containing the conditional information – those who lose their jobs because of a departmental reorganization.

Characterizing document literacy tasks

There are 54 tasks ordered along the 500-point document literacy scale. These 54 tasks comprise 19 items from IALS and 35 new tasks developed for ALL. Together, these tasks range in difficulty value from 157 to 444. A Level 1 document literacy task with a difficulty value of 188 directs the reader to identify from a chart the percentage of teachers from Greece who are women. The chart shown here displays the percentage of teachers from various countries who are women. In terms of our process variables, type of match was judged to be easy because the reader was required to locate a single piece of information that was literally stated in the chart; type of information was judged to be relatively easy because it was an amount; and plausibility of distractor is also judged to be relatively easy because there are distractors for the requested information.

FEW DUTCH WOMEN AT THE BLACKBOARD

There is a low percentage of women teachers in the Netherlands compared to other countries. In most of the other countries, the majority of teachers are women. However, if we include the figures for inspectors and school principals, the proportion shrinks considerably and women are in a minority everywhere.

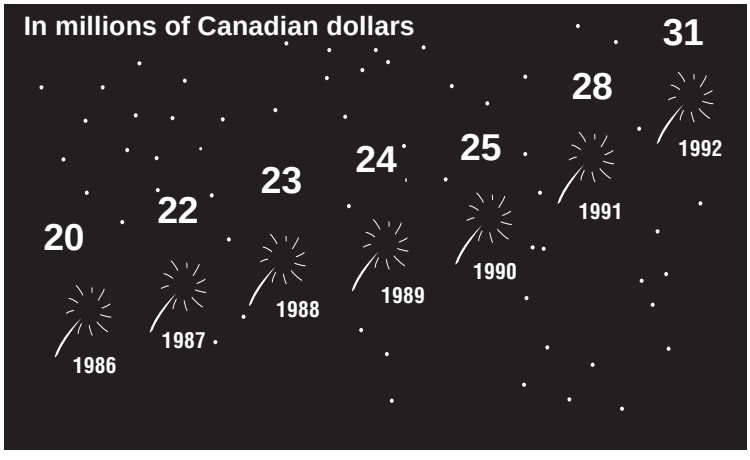


Percentage of women teachers (kindergarten, elementary, and secondary).

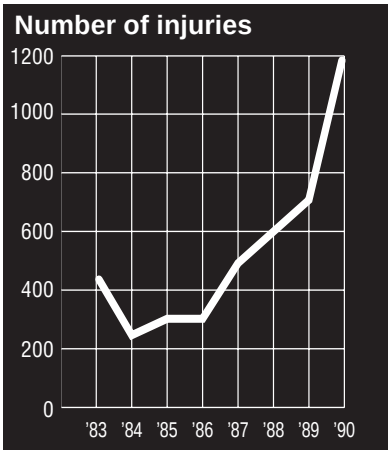
A second document task involving this same chart directs the reader to identify the country other than the Netherlands in which women teachers are in the minority. This item falls in the middle of Level 2 and received a difficulty value of 234. This task was made a bit more difficult than the first because rather than searching for a country and locating a percentage, the reader had to know that minority means less than 50 per cent. Then they had to cycle through to identify the countries in which the percentage of women teachers were less then 50 per cent. In addition, they had to remember the condition "other than the Netherlands"; otherwise they might have chosen it over the correct response. As a result, type of match was scored as moderately difficult; type of information as easy because the requested information is a country or place; and plausibility of distractor as relatively easy because there are distractors associated with the requested information.

A somewhat more difficult task, with a difficulty value of 295 and falling in the middle of Level 3 directs the reader to look at charts involving fireworks from the Netherlands and to write a brief description of the relationship between sales and injuries based on the information shown. Here the reader needs to look at and compare the information contained in the two charts and integrate this information making an inference regarding the relationship between the two sets of information. As a result, it was judged as being relatively difficult in terms of type of match. Type of information also was judged to be relatively difficult because the requested information is asking for a pattern or similarity in the data. Plausibility of distractor was scored moderately difficult primarily because both given and requested information is present in the task. For example, one of the things that may have contributed to the difficulty of this task is the fact that the sales graph goes from 1986 to 1992 while the injuries graph goes from 1983 to 1990. The reader needed to compare the information from the two charts for the comparable period time.

Fireworks in the Netherlands



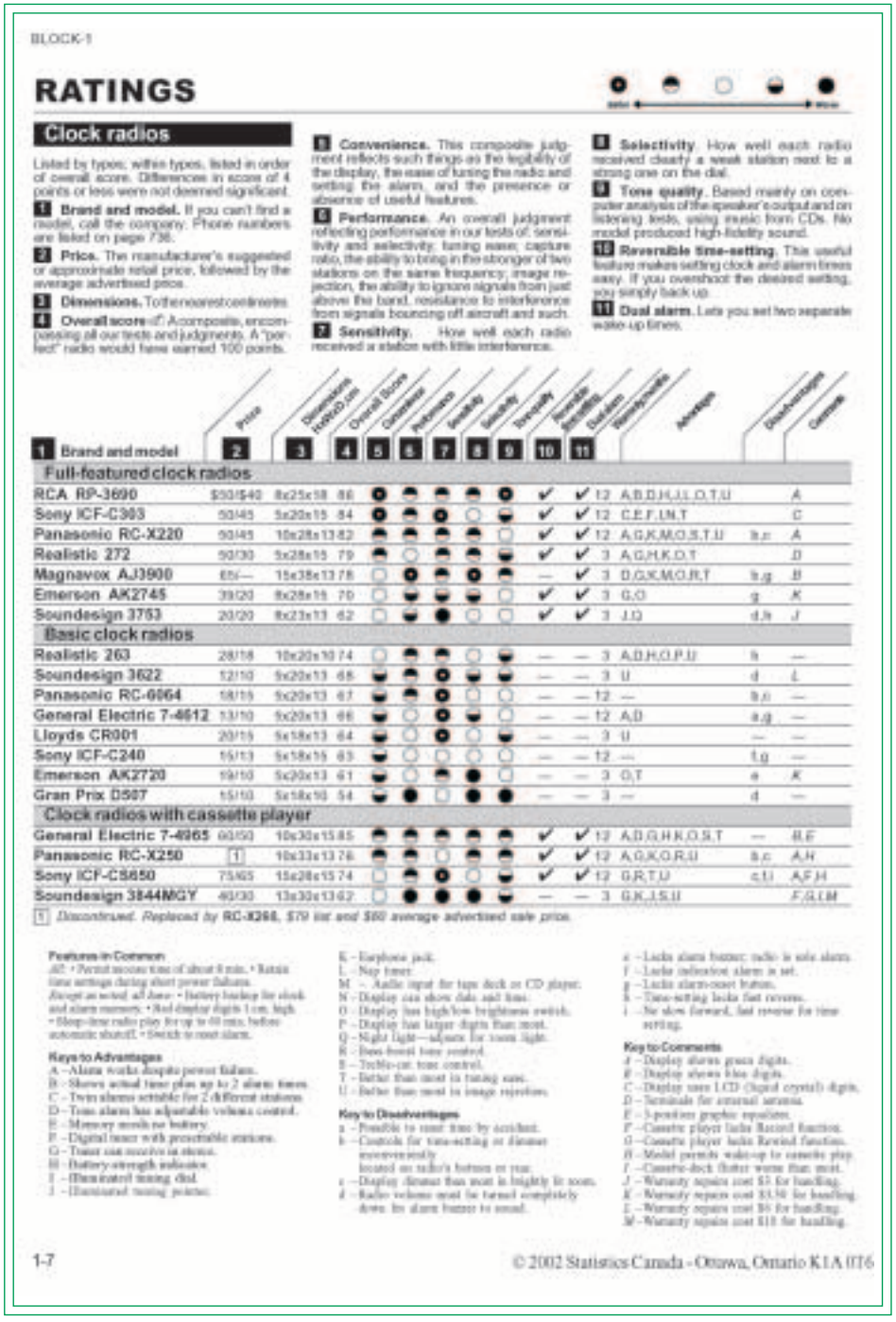
Victims of fireworks



Another set of tasks covering a range of difficulty on the document scale involved a rather complicated document taken from a page in a consumer magazine rating clock radios. The easiest of the three tasks, receiving a difficulty value of 287 and falling in Level 3, asks the reader "which two features are not on any basic clock radio." In looking at the document, the reader has to cycle through the document, find the listing for basic clock radios, and then determine that a dash represents the absence of a feature. They then have to locate the two features indicated by the set of dashes. As a result, type of match was judged as being relatively difficult because it is a cycle requiring multiple responses with a condition or low text based inference. Type of information was scored as relatively easy because its features are an attribute of the clock radio and plausibility of distractor is relatively easy because there are some characteristics that are not associated with other clock radios.

A somewhat more difficult task associated with this document and falling in the lower end of Level 4 received a difficulty value of 327. It asks the reader "which full-featured clock radio is rated highest on performance." Here the reader must make a three-feature match (full-featured, performance, and highest) where one of the features requires them to process conditional information. It is possible, for example, that some readers were able to find the full-featured radios and the column listed under performance but selected the first radio listed assuming it was the one rated highest. In this case, they did not understand the conditional information which is a legend stating what the symbols mean. Others may have gone to the column labelled overall score and found the highest numerical number and chosen the radio associated with it. For this reason, plausibility of distractor was scored as moderately difficult. Type of information was judged as being easy because the requested information is a thing.

The most difficult task associated with this document, with a difficulty level of 408, and falling in Level 5 asks the reader to identify the average advertised price for the basic clock radio receiving the highest overall score. This task was made more difficult because the reader had to match four rather than three features; they also had to process conditional information and there was a highly plausible distractor in the same node as the correct answer. As a result of these factors, type of match was judged to be relatively difficult, type of information relatively easy and plausibility of distractor as having the highest level of difficulty.



BLOCK-1

RATINGS

Clock radios

Listed by types, within types, listed in order of overall score. Differences in score of 4 points or less were not deemed significant.

1 Brand and model. If you can't find a model, call the company. Phone numbers are listed on page 736.

2 Price. The manufacturer's suggested or approximate retail price, followed by the average advertised price.

3 Dimensions. To the nearest centimetre.

4 Overall score (0 to 100). A composite, encompassing all our tests and judgments. A "perfect" radio would have earned 100 points.

5 Convenience. This composite judgment reflects such things as the legibility of the display, the ease of tuning the radio and setting the alarm, and the presence or absence of useful features.

6 Performance. An overall judgment reflecting performance in our tests of: sensitivity and selectivity; tuning ease; capture ratio, the ability to bring in the stronger of two stations on the same frequency; image rejection, the ability to ignore signals from just above the band; resistance to interference from signals bouncing off aircraft and such.

7 Sensitivity. How well each radio received a station with little interference.

8 Selectivity. How well each radio received clearly a weak station next to a strong one on the dial.

9 Tone quality. Based mainly on computer analysis of the speaker's output and on listening tests, using music from CDs. No model produced high-fidelity sound.

10 Reversible time-setting. This useful feature makes setting clock and alarm times easy. If you overshoot the desired setting, you simply back up.

11 Dual alarm. Lets you set two separate wake-up times.

1 Brand and model	Price 2	Dimensions H×W×D cm 3	Overall Score 4	Convenience 5	Performance 6	Sensitivity 7	Selectivity 8	Tone quality 9	Reversible time-setting 10	Dual alarm 11	Warranty (months)	Advantages	Disadvantages	Comments
Full-featured clock radios														
RCA RP-3690	$30/$40	8x25x18	86	●	◓	◓	◓	●	✔	✔	12	A,B,D,H,J,L,O,T,U		A
Sony ICF-C303	50/45	5x20x15	84	●	◓	○	○	◒	✔	✔	12	C,E,F,I,N,T		C
Panasonic RC-X220	50/45	10x28x13	82	◓	◓	◓	◓	□	✔	✔	12	A,G,K,M,O,S,T,U	b,c	A
Realistic 272	50/30	5x28x15	79	◓	○	◓	◓	◒	✔	✔	3	A,G,H,K,O,T		D
Magnavox AJ3900	65/—	15x38x13	78	□	●	◓	●	◓	—	✔	3	D,G,K,M,O,R,T	b,g	B
Emerson AK2745	39/20	8x28x15	70	□	◒	◒	◒	□	✔	✔	3	G,O	g	K
Soundesign 3753	20/20	8x23x13	62	□	◒	●	□	□	✔	✔	3	J,Q	d,h	J
Basic clock radios														
Realistic 263	28/18	10x20x10	74	□	◓	◓	□	◒	—	—	3	A,D,H,O,P,U	h	—
Soundesign 3622	12/10	5x20x13	68	◒	◓	●	◒	◒	—	—	3	U	d	L
Panasonic RC-6064	18/15	5x20x13	67	◒	◓	●	□	□	—	—	12	—	b,c	—
General Electric 7-4612	13/10	5x20x13	66	◒	○	●	◒	—	—	—	12	A,D	a,g	—
Lloyds CR001	20/15	5x18x13	64	◒	○	●	□	◒	—	—	3	U	—	—
Sony ICF-C240	15/13	5x18x15	63	◒	○	□	□	□	—	—	12	—	f,g	—
Emerson AK2720	19/10	5x20x13	61	◒	○	◓	●	□	—	—	3	O,T	a	K
Gran Prix D507	15/10	5x18x10	54	◒	●	□	●	●	—	—	3	—	d	—
Clock radios with cassette player														
General Electric 7-4965	60/50	10x30x15	85	◓	◓	◓	◓	◓	✔	✔	12	A,D,G,H,K,O,S,T	—	B,E
Panasonic RC-X250	[1]	10x33x13	76	◓	◓	□	◓	◓	✔	✔	12	A,G,K,O,R,U	b,c	A,H
Sony ICF-CS650	75/65	15x28x15	74	□	◓	●	□	◒	✔	✔	12	G,R,T,U	c,i	A,F,H
Soundesign 3844MGY	40/30	13x30x13	62	□	●	●	●	◒	—	—	3	G,K,J,S,U		F,G,I,M

[1] *Discontinued. Replaced by* **RC-X266**, *$79 list and $60 average advertised sale price.*

Features in Common
All: • Permit snooze time of about 8 min. • Retain time settings during short power failures. *Except as noted, all have:* • Battery backup for clock and alarm memory. • Red display digits 1 cm. high. • Sleep-time radio play for up to 60 min. before automatic shutoff. • Switch to tone alarm.

Keys to Advantages
A – Alarm works despite power failure.
B – Shows actual time plus up to 2 alarm times.
C – Two alarms settable for 2 different stations.
D – Tone alarm has adjustable volume control.
E – Memory needs no battery.
F – Digital tuner with presettable stations.
G – Tuner can receive in stereo.
H – Battery-strength indicator.
I – Illuminated tuning dial.
J – Illuminated tuning pointer.
K – Earphone jack.
L – Nap timer.
M – Audio input for tape deck or CD player.
N – Display can show date and time.
O – Display has high/low brightness switch.
P – Display has larger digits than most.
Q – Night light – adjusts for room light.
R – Bass-boost tone control.
S – Treble-cut tone control.
T – Softer than most in tuning ease.
U – Softer than most in image rejection.

Keys to Disadvantages
a – Possible to reset time by accident.
b – Controls for time-setting or dimmer inconveniently located on radio's bottom or rear.
c – Display dimmer than most in brightly lit room.
d – Radio volume must be turned completely down for alarm buzzer to sound.
e – Lacks alarm buzzer; radio is sole alarm.
f – Lacks indicator alarm is set.
g – Lacks alarm-reset button.
h – Time-setting locks fast reverse.
i – No slow forward, fast reverse for time setting.

Key to Comments
A – Display shows green digits.
B – Display shows blue digits.
C – Display uses LCD (liquid crystal) digits.
D – Terminals for external antenna.
E – 5-position graphic equalizer.
F – Cassette player lacks Record function.
G – Cassette player lacks Rewind function.
H – Model permits wake-up to cassette play.
I – Cassette-deck flutter worse than most.
J – Warranty repairs cost $5 for handling.
K – Warranty repairs cost $3.50 for handling.
L – Warranty repairs cost $6 for handling.
M – Warranty repairs cost $10 for handling.

1-7

Measuring numeracy in ALL

Defining numeracy in ALL

The conception of numeracy developed for ALL is built upon recent research and work done in several countries on functional demands of different life contexts, on the nature of adults' mathematical and statistical knowledge and skills, and on how such skills are applied or used in different circumstances. In light of the general intention of the ALL survey to provide information about a diverse set of life skills, this framework defines numeracy as follows:

Numeracy is the knowledge and skills required to effectively manage and respond to the mathematical demands of diverse situations.

This definition implies that numeracy is broader than the construct of quantitative literacy defined by IALS[2]. Further, adult numeracy should be viewed as different from "knowing school mathematics". Although a universally accepted definition of "numeracy" does not exist (Baker and Street, 1994), an examination of some perspectives on the meaning of adult numeracy shows that they contain many commonalities. Below are two examples, both from work in Australia:

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).

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 (Queensland Department of Education, 1994)

All these definitions are quite similar, in their broad scope, to the ALL definitions of prose and document literacy presented in a prior section. Many conceptions of numeracy emphasize the practical or functional application and use of mathematical knowledge and skills to cope with the presence of mathematical elements in real situations. Adults are expected to possess multiple ways of responding flexibly to a mathematical situation in a goal-oriented way, dependent on the needs and interests of the individual within the given context (i.e., home, community, workplace, etc...), as well as on his or her attitudes and beliefs toward numeracy (Gal, 2000; Coben, O'Donoghue and FitzSimons, 2000).

Thus, numeracy involves more than just applying arithmetical skills to information embedded in printed materials, which was the focus of assessment in IALS. Adult numeracy extends to a possession of number sense, estimation skills, measurement and statistical literacy. Given the extent to which numeracy pervades the modern world, it is not necessarily just commonly encountered situations that require numerate behaviour, but also *new* situations.

Another important element in defining numeracy is the role of communication processes. 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 textual or symbolic messages as well as to communicate mathematical information and reasoning processes (Marr and Tout, 1997; Gal, 1997).

Definitions of numeracy explicitly state that numeracy not only refers to operating with numbers, as the word can suggest, especially to those familiar with conceptions of children's numeracy, but covers a wide range of mathematical skills and understandings. Further, in recent years there has been much discussion and debate about the relationship between mathematics and numeracy and about the concept of "critical" numeracy (Frankenstein, 1989; Steen, 2001). Johnston, 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).

Many authors argue that a discussion of functional skills should also address supporting or enabling attitudes and beliefs. In the area of adults' mathematical skills, "at homeness" with numbers or "confidence" with mathematical skills is expected, as these affect how skills and knowledge are actually put into practice (Cockroft, 1982; Tobias, 1993).

The brief definition of numeracy developed for ALL and presented earlier above is complemented by a broader definition of *numerate behaviour* which was developed by the ALL Numeracy Team to serve as the basis for the development of numeracy items for ALL:

Numerate behaviour 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, factors and processes.

This conception of numerate behaviour implies that in order to assess people's numeracy, it is necessary to generate tasks and items which vary in terms of contexts, the responses called for, the nature of the mathematical information involved, and the representations of this information. These task characteristics are elaborated below. This conception is much broader than the definition of quantitative literacy used in IALS. Its key elements relate in a broad way to situation management and to a need for a range of responses (not only to responses that involve numbers). It refers to a wide range of skills and knowledge (not only to application of arithmetical knowledge and computational operations) and to the use of a wide range of situations that present respondents with mathematical information of different types (not only those involving *numbers* embedded in *printed* materials).

The item development process aimed to ensure that a certain proportion of the item pool would place a minimum reading burden on the respondents, i.e., that some of the stimuli would be text-free or almost so, allowing even respondents with limited mastery of the language of the test to comprehend the situation described. Other parts of the item pool included items requiring varying amounts of essential texts as dictated by the situation which the item aimed to represent.

As implied by the literature and ideas reviewed earlier, the nature of a person's responses to the mathematical and other demands of a situation will depend critically on the activation of various enabling knowledge bases (understanding of the context; knowledge and skills in the areas of mathematics, statistics and literacy), on reasoning processes and on their attitudes and beliefs with respect to numeracy. In addition, numerate behaviour requires the integration of mathematical knowledge and skills with broader literacy and problem solving skills along with the prior experiences and practices that each person brings to every situation. It is clear that numerate behaviour will involve an attempt to engage with a task and not delegate it to others or deal with it by intentionally ignoring its mathematical content.

Identifying task characteristics

Four key characteristics of numerate behaviour were used to develop and represent the numeracy tasks built for ALL – type of purpose/context, type of response, type of mathematical or statistical information, and type of representation of mathematical or statistical information. Each of these is described next.

Type of purpose/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 those that one faces in personal and family life, or revolve around hobbies, personal development, or interests. Representative tasks are handling money and budgets, comparison shopping, planning nutrition, personal time management, making decisions involving travel, planning trips, 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 have to develop skills in managing situations that might be narrower in their application of mathematical themes. Representative tasks are completing purchase orders, totalling receipts, calculating change, managing schedules, using spreadsheets, organizing and packing different shaped goods, completing and interpreting control charts or quality graphs, making and recording measurements, reading blueprints, tracking expenditures, predicting costs and applying formulas.

Societal or community

Adults need to know about processes happening in the world around them, such as trends in crime, wages and employment, pollution, medical or environmental risks. They may have to take part in social or community events, or in political action. This requires that adults can read and interpret quantitative information presented in the media, including statistical messages and graphs. They may have to manage situations like organizing a fund-raiser, planning fiscal aspects of a community program, or interpreting the results of a study about risks of the latest health fad.

Further learning

Numeracy skills 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 know some of the more formal aspects of mathematics that involve symbols, rules, and formulas and to understand some of the conventions used to apply mathematical rules and principles.

Type of responses. 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):

Identify or locate some mathematical information present in the task or situation confronting them that is relevant to their purpose or goal.

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).

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 judgment about how mathematical information or known facts actually apply to the situation or context. Contextual judgment may have to be used in deciding whether an answer makes sense or not in the given context, for example, that a result of "2.35 cars" is not a valid solution to how many cars are needed to transport a group. It can also incorporate a critical aspect, where a person questions the purpose of the task, the validity of the data or information presented, and the meaning and implications of the results, both for them as an individual and possibly for the wider community.

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 (ranging from a simple number or word to a detailed explanation or analysis) and/or through drawing (a diagram, map, graph).

Type of mathematical or statistical information. 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 and Ahlgren (1990) described networks of related ideas: numbers, shapes, uncertainty, summarizing data, sampling and reasoning. Dossey (1997) categorized the mathematical behaviours of quantitative literacy as: data representation and interpretation, number and operation sense, measurement, variables and relations, geometric shapes and spatial visualization, and chance. The ALL Numeracy Team drew from these and other closely tied categorizations (e.g., National Council of Teachers of Mathematics, 2000) to arrive at a set of five fundamental ideas that characterize the mathematical demands facing adults in diverse situations at the beginning of the 21st century.

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 per cents as expressions of greater precision, or as indications of parts-of-whole which are useful when comparing proportions.

Positive and negative numbers serve 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" for magnitude and the meaning of very large or very small numbers, and sometimes a sense for the relative magnitude of different proportions.

Money and time management, the ubiquitous mathematics that is part of every adult's life, depends on a good sense of number and quantity. Contextual judgment comes into play when deciding how precise one should be when conducting certain computations or affects the choice of which tool (calculator, mental math, a computer) to use. A low level numeracy task might be figuring out the cost of one can of soup, given the cost of four for $2.00; a task with a higher cognitive demand could involve "harder numbers" such as when figuring out the cost per kilo while buying 0.783 kg of cheese for 12,95 Euros.

Dimension and shape

Dimension includes "big ideas" related to one, two and three dimensions of "things". Understanding of dimensions is called for when encountering or generating spatial or numerical descriptions of objects, making projections, or working with lengths, perimeters, planes, surfaces, location, etc… Facility with each dimension requires a sense of "benchmark" measures, direct measurement, and estimations of measurements.

Shape is a category describing real or imaginary images and entities that can be visualized (e.g., houses and buildings, designs in art and craft, safety signs, packaging, knots, crystals, shadows and plants). 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 more complex task might involve describing the change in the size or volume of an object when one dimension is changed, such as when choosing between different boxes for packaging certain objects.

Pattern, functions 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, nature, traffic patterns, etc… 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 analytically underlies 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 functions, relationships between variables, is a crucial gateway to understanding even the most basic economic, political or social analyses. A relatively simple pattern-recognition task might require someone to describe the pattern in a sequence of given numbers or shapes, and in a functional context to understand the relationship between lists or variables (e.g., weight and volume of objects); having to develop a formula for an electronic spreadsheet would put a higher level of demand on the individual.

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 analysis, and common measures of center or spread, or the idea of a statistical inference. Modern society demands that adults are able to interpret (and at times even produce) frequency tables, basic charts and graphs, information about averages and medians, as well as identify questionable statistical claims (Gal, 2002).

Chance covers "big ideas" related to probability and relevant statistical concepts and tools. Few things in the world are 100 per cent certain; thus the ability to attach a number that represents the likelihood of an event (including risks or side-effects) is a valuable tool whether it has to do with the weather, the stock-market, or the decision to use a certain drug. In this category, a simple numeracy skill might be the interpretation of a simple pie chart or comprehension of a statement about an average; a more complex task would be to infer the likelihood of occurrence of an event based upon given information.

Change

This term 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 or exponential patterns of change can characterize steady trends; periodic changes suggest cycles and irregular change patterns connect with chaos theory. Describing weight loss over time is a relatively simple task, while calculating compounded interest is a relatively complex task.

Type of 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., sheep, people, buildings, cars, etc…) or as pictures of such things. It may be conveyed through symbolic notation (e.g., numerals, letters, or operation signs). Sometimes, mathematical information will be conveyed by formulas, which are a model of relationships between entities or variables.

Further, mathematical information may be encoded in visual displays such as *diagrams* or *charts*; *graphs*, and *tables* may be used to display aggregate statistical or quantitative information. Similarly, *maps* of real entities (e.g., of a city or a project plan) may contain numerical data but also information that can be quantified or mathematized.

Finally, a person may have to extract mathematical information from various types of texts, either in prose or in documents with specific formats (such as in tax forms). Two different kinds of text may be encountered in functional numeracy tasks. The first involves mathematical information represented in textual form, i.e., with words or phrases that carry mathematical meaning. Examples are the use of number words (e.g., "five" instead of "5"), basic mathematical terms (e.g., fraction, multiplication, per cent, average, proportion), or more complex phrases (e.g., "crime rate cut by half") that require interpretation. The second involves cases where mathematical information is expressed in regular notations or symbols (e.g., numbers, plus or minus signs, symbols for units of measure, etc…), but is surrounded by text that despite its non-mathematical nature also has to be

interpreted in order to provide additional information and context. An example is a bank deposit slip with some text and instructions in which numbers describing monetary amounts are embedded.

Characterizing numeracy tasks

A total of 40 numeracy tasks were selected and used in the ALL survey. These tasks range along the numeracy scale from 174 to 380 and their placement was determined by how well adults in participating countries responded to each task. Described below are sample tasks that reflect some of the conceptual facets of the numeracy construct and scale design principles described earlier, such as computations, spatial and proportional reasoning, measurement, and statistical knowledge.

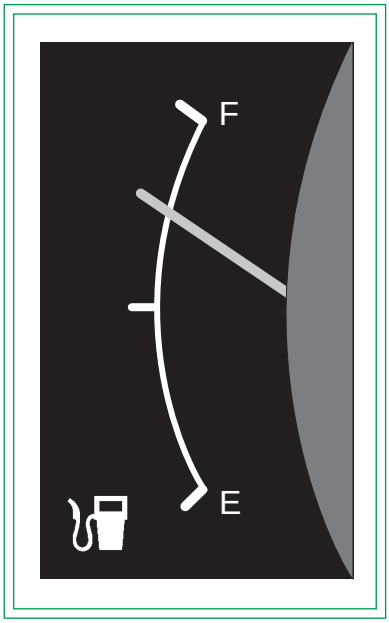
As expected, the easiest task on the numeracy scale required adults to look at a photograph containing two cartons of coca cola bottles (174). They were directed to find the total number of bottles in the two full cases being shown. Part of what made this task easy is the fact that content was drawn from everyday life and objects of this kind would be relatively familiar to most people. Second, what adults were asked to do was apparent and explicit – this tasked used a photograph depicting concrete objects and required the processing of no text. A third contributing factor is that respondents could approach the task in a variety of ways that differ in sophistication, such as by multiplying rows and columns, but also by simple counting. This task requires that adults make a conjecture since the full set of bottles in the lower case is not visible, but as can be seen from the low difficulty level of the task, this feature did not present a problem for the vast majority of adults in all participating countries.

A second task that was also quite easy directed adults to look at a short text depicting the results of an election involving three candidates and determine the total number of votes cast. This task received a difficulty value of 192, falling in Level 1 on the numeracy scale. Again, respondents were asked to deal with a realistic type of situation where simple numerical information is displayed in a

simple column format showing the name of each candidate and the number of votes that the candidate received. No other numerical information was present that can be a distractor. Finding the total number of votes cast in the election requires a single addition operation that is made explicit in the question by the use of the keyword "total", and the computation involves relatively small whole numbers.

A more complex numeracy task falling in the middle of Level 2 and receiving a difficulty value of 248 directs adults to look at a gas (petrol) gauge. This gauge has three lines or ticks on it with one showing an "F", one showing an "E" and the third in the middle between the two. A line on the gauge, representing the gauge's needle, shows a level that is roughly halfway between the middle tick and the tick indicating "F", suggesting that the tank is about three-quarters full. The directive states that the tank holds 48 gallons and asks the respondent to determine "how many gallons remain in the tank." This task is drawn from an everyday context and requires an adult to interpret a display that conveys quantitative information but carries virtually no text or numbers. No mathematical information is present other than what is given in the question.

What makes this task more difficult than the previous ones described above is the fact that adults must first estimate the level of gas remaining in the tank, by converting the placement of the needle to a fraction. Then they need to determine how many gallons this represents from the 48 gallon capacity stated in the question or directive. Thus, this task requires adults to apply multiple operations or procedures to arrive at a correct response, without specifying what the operations may be. Nonetheless, this task, like many everyday numeracy tasks, does not require an exact computation but allows an approximation that should fall within reasonable boundaries.



A somewhat more difficult numeracy task, falling at the top of Level 2 and receiving a difficulty value of 275, requires adults to look at a diagram of a container on which there are four markings or lines; respondents are asked to draw a line on the container indicating where one-third would be. The top line is marked "1" while the middle line is marked with "1/2". There are two other lines with no markings - one line midway between "1" and "1/2" and another midway between

the line marked "1/2" and the bottom of the container. To respond correctly, adults need to mark a line on the container that is between the line marked "1/2" and the line below it indicating where one-quarter would be (although this line does not say "1/4" – this has to be inferred). Here the context may be less familiar to the respondent but again the visual image used is simple and realistic with virtually no text; the response expected does not involve writing a symbol or text, just drawing a line in a certain region on the drawing of the container. To answer this task correctly, adults need to have some working knowledge of fractions and a sense for proportions: they have to be familiar with the symbols for "1/2" and "1/3", know how to order fractions in terms of their relative size and be able to relate them to the existing markings on the container.

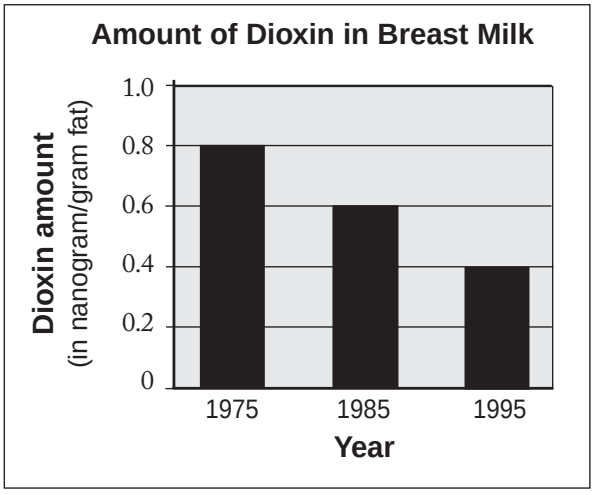
Some numeracy tasks were developed around a short newspaper article titled "Is breast milk safe?" which relates to environmental hazards and food safety. The article contained two brief text paragraphs describing a toxin, Dioxin, found in fish in the Baltic Sea plus a graph with bars indicating the levels of Dioxin found at three points in time, namely 1975, 1985, and 1995, in the breast milk of North European women. One question asked adults to describe how the amount of Dioxin changed from 1975 to 1995, i.e., provide a straightforward interpretation of data presented in a graph. Adults were not required to actually calculate the amount of change over each of the periods, just describe in their own words the change in the levels of Dioxin (e.g., decreased, increased, stayed the same).

This task received a difficulty value of 280, the lower end of Level 3. The graph clearly indicates that the amount of Dioxin decreased over each of the three time periods, yet some adults have difficulty coping with such a task, which is based on a stimulus with a structure that commonly appears in newspapers, i.e., brief text plus a graph. The increased difficulty level of this item may be attributable in part to the need for adults to generate their own description, to the moderate amount of dependence on text needed to comprehend the context to which the graph refers, or to the need to understand the direction of the decimal values on the vertical axis (which is common in reporting on concentrations of contaminating chemicals).

Is breast milk safe?

Since the 1970s, scientists have been worried about the amount of Dioxin, a toxin in fish caught in the Baltic sea. Dioxin tends to accumulate in breast milk and can harm newborn babies.

The diagram shows the amount of Dioxin in the breast milk of North European women, as found in studies done from 1975 to 1995.



A second and more difficult task using this same stimulus directed adults to compare the per cent of change in Dioxin level from 1975 to 1985 to the per cent of change in Dioxin level from 1985 to 1995, determine which per cent of change is larger, and explain their answer. This task was considerably more difficult for adults in participating countries and received a difficulty value of 377 on the numeracy scale. Here the necessary information is embedded within the graph and requires a level of transformation and interpretation. To arrive at a correct response, adults have to look at the rate of change expressed in per cents, not just the absolute size of the change. Further, they have to work with per cents of entities smaller than one (i.e., the decimal values on the vertical axis) and realize that the base for the computation of per cent change shifts for each pair. It seems that the need to cope with such task features, use formal mathematical procedures, or deal with the abstract notion of rate of change, adds considerable difficulty to such tasks.

The most difficult numeracy task in this assessment, receiving a difficulty value of 380 (Level 5), presented adults with an advertisement claiming that it is possible for an investor to double an amount invested in seven years, based on a 10 per cent fixed interest rate each year. Adults were asked if it is possible to double $1000 invested at this rate after seven years and had to support their answer with their calculations. A range of responses was accepted as correct as long as a reasonable justification was provided, with relevant computations. Respondents were free to perform the calculation any way they wanted, but could also use a "financial hint" which accompanied the advertisement and presented a formula for estimating the worth of an investment after any number of years. Those who used the formula had to enter information stated in the text into variables in the formula (principal, interest rate and time period) and then perform the needed computations and compare the result to the expected amount if $1000 is doubled.

All respondents could use a hand-held calculator provided as part of the assessment. This task proved difficult because it involved per cents and the computation, whether with or without the formula, required the integration of several steps and several types of operations. Performing the computations without the formula required understanding of compound interest procedures. This task allowed adults to use a range of reasoning strategies, including informal or invented procedures. Yet, like the previous task involving the comparison of rates of change, it required the use of formal mathematical information and deeper understanding of non-routine computational procedures, all of which may not be familiar or accessible to many adults.

Measuring problem solving in ALL

Defining problem solving in ALL

Research on problem solving has a long tradition within both academic psychology and applied human resources research. A very general definition of problem solving that reflects how it is generally understood in the psychological literature (Hunt, 1994; Mayer, 1992; Mayer and Wittrock, 1996; Smith, 1991) is presented here:

> *Problem solving is goal-directed thinking and action in situations for which no routine solution procedure is available.* **The problem solver has a more or less well-defined goal, but does not immediately know how to reach it. The incongruence of goals and admissible operators constitutes a problem. The understanding of the problem situation and its step-by-step transformation, based on planning and reasoning, constitute the process of problem solving.**

One major challenge while developing a framework for problem solving that is to be used in a survey such as ALL is how best to adapt the psychological literature to the constraints imposed by a large-scale international comparative study. In order to do this, a decision was made to focus on an essential subset of problem solving – analytical problem solving. Our notion of analytical problem solving is not to be confused with the intuitive everyday use of the term or with the clinical-psychological concept in which problem solving is associated with the resolution of social and emotional conflicts. Nevertheless, social context is also relevant for our definition of analytical problem solving, for example when problems have to be approached interactively and resolved through co-operation. Motivational factors such as interest in the topic and task-orientation also influence the problem-solving process. However, the quality of problem solving is primarily determined by the comprehension of the problem situation, the thinking processes used to approach the problem, and the appropriateness of the solution.

The *problem* itself can be characterized by different aspects:

- The *context* can reflect different domains, which may be of a theoretical or a practical nature, related to academic situations or to the real world. Within these domains, problems can be more or less authentic.
- The *scope* of a problem can range from working on limited, concrete parts of a task to planning and executing complex actions or evaluating multiple sequences of actions.
- The problem can have a well-defined or an ill-defined goal, it can have transparent (explicitly named) or non-transparent constraints, and involve few independent elements or numerous interconnected ones. These features determine the *complexity* of the problem.

How familiar the context is to the target population, whether the problem involves concrete tasks or complex actions, how well the goal is defined, how transparent the constraints are, how many elements the problem solver has to take into account and how strongly they are interconnected – are all features that will determine the level of problem-solving competency required to solve a certain

problem. The empirical difficulty, i.e., the probability of giving a correct solution, will depend on the relation between these problem features on the one hand, and the subjects' competency level on the other hand.

The *cognitive processes* that are activated in the course of problem solving are diverse and complex, and they are likely to be organized in a non-linear manner. Among these processes, the following five components may be identified:

1. Searching for information, and structuring and integrating it into a mental representation of the problem ("situational model").
2. Reasoning, based on the situational model.
3. Planning actions and other solution steps.
4. Executing and evaluating solution steps.
5. Continuous processing of external information and feedback.

Baxter and Glaser (1997) present a similar list of cognitive activities labelled "general components of competence in problem solving": problem representation, solution strategies, self-monitoring, and explanations. Analytical problem solving in everyday contexts, as measured by the ALL problem-solving instrument, focuses on the components 1 to 3 listed above (and to some extent 4).

One of the most important insights of recent research in cognitive psychology is that solving demanding problems requires at least some knowledge of the domain in question. The concept of a problem space through which a General Problem Solver moves by means of domain-independent search strategies (Newell and Simon, 1972) proved to be too simple to describe how problem situations are understood and the process of finding a solution. Efforts to identify a general, domain-independent competence for steering dynamic systems (operative intelligence) within the framework of complex problem-solving research were also unsuccessful; performance on such systems can only partially be transferred to other systems (Funke, 1991). However, research on grade 3 to grade 12 students showed that problem-solving skills clearly improve under well-tuned training conditions and that a substantial transfer across different problems can be achieved (Reeff et al. 1989, 1992, 1993; Regenwetter, 1992; Regenwetter and Müller, 1992; Stirner, 1993).

Problem solving is dependent on knowledge of concepts and facts (declarative knowledge) and knowledge of rules and strategies (procedural knowledge) in a given subject domain. Although it is evident from past research that declarative knowledge in the problem domain can substantially contribute to successful problem-solving strategies, procedural knowledge is crucial as well. The amount of relevant previous knowledge available could also account for the relation between intelligence and problem-solving performance, as shown in the work of Raaheim (1988) and Leutner (1999). People with no relevant previous knowledge at all are unable to explore the problem situation or plan a solution in a systematic manner and are forced to rely on trial and error instead. Those who are already very familiar with the task are able to deal with it as a matter of routine. General intellectual ability, as measured by reasoning tasks, plays no role in either of these cases. When problem solvers are moderately familiar with the task, analytical reasoning strategies can be successfully implemented.

The approach taken for the assessment of problem solving in ALL relies on the notion of (moderately) familiar tasks. Within a somewhat familiar context the problems to be solved are inexplicit enough so as not to be perceived as pure routine tasks. On the other hand, the domain-specific knowledge prerequisites are sufficiently limited as to make analytical reasoning techniques the main cognitive tool for solving the problems.

Identifying task characteristics

How can contextualized, real-life problems be defined and transformed into a set of assessment tasks? After reviewing the various approaches that have been taken in previous research to measure problem solving, a decision was made to use a project approach in ALL. The project approach has the potential to be a powerful means for assessing analytical problem solving skills in real world, everyday contexts for several reasons. Solving problems in project-like settings is important and relevant for adults in both their professional and their private life. In addition, the project approach has been successfully implemented in other large-scale assessments, and it can be realized as a paper-and-pencil-instrument, which is of crucial importance for contemporary large-scale surveys. Furthermore, the project approach uses different problem-solving stages as a dimension along which to generate the actual test items. Following Pólya (1945, 1980), the process of problem solving has been frequently described in terms of the following stages:

- Define the goal.
- Analyze the given situation and construct a mental representation.
- Devise a strategy and plan the steps to be taken.
- Execute the plan, including control and – if necessary – modification of the strategy.
- Evaluate the result.

The different action steps define the course of action for an "everyday" project. One or more tasks or items are generated to correspond to each of these action steps. Respondents are expected to work on individual tasks that have been identified as steps that need to be carried out as a part of their project (a sample project, for example, might involve "planning a reunion" or "renovating a clubhouse"). Embedding the individual tasks in a project is believed to yield a high degree of context authenticity. Although they are part of a comprehensive and coherent project, the individual tasks are designed so that they can be solved independently of one another and are expected to vary in complexity and overall difficulty for adults.

Since assessing problem solving skills in large-scale assessments is a relatively new endeavour, it might be helpful to provide a detailed account of the construction process. Table A1 provides an overview of the problem solving steps as they correspond to the action steps identified above. Different components and aspects of each of the problem solving steps are listed.

TABLE A1

Problem-solving steps and instantiations

Define the goals	• Set goals. • Recognize which goals are to be reached and specify the essential reasons for the decision. • Recognize which goals/wishes are contradictory and which are compatible. • Assign priorities to goals/wishes.
Analyze the situation	• Select, obtain and evaluate information. ⇒ What information is required, what is already available, what is still missing, and what is superfluous? ⇒ Where and how can you obtain the information? ⇒ How should you interpret the information? • Identify the people (e.g. with what knowledge and skills) who are to be involved in solving the problem. • Select the tools to be used. • Recognize conditions (e.g. time restrictions) that need to be taken into account.
Plan the solution	• Recognize which steps need to be taken. • Decide on the sequence of steps (e.g. items on the agenda). • Coordinate work and deadlines. • Make a comparative analysis of alternative plans (recognize which plan is suitable for reaching the goals). • Adapt the plan to changed conditions. • Opt for a plan.
Execute the plan	• Carry out the individual steps (e.g., write a letter, fill in a form, make calculations).
Evaluate the results	• Assess whether and to what extent the target has been reached. • Recognize mistakes. • Identify reasons for mistakes. • Assess consequences of mistakes.

The construction of a pool of assessment tasks that could be mapped back to these five action steps involved several phases of activities. First was the identification of appropriate projects that would be suitable for adults with varying educational backgrounds and relevant to the greatest number of people in the target group. Next, developers had to identify and sketch out the problem situation and the sequence of action steps that relate back to the model. Third, they had to develop a pool of items that were consistent with the action steps and that tapped into particular processes including the development of correct responses and appropriate distractors for multiple choice items and solution keys and scoring guides for open-ended tasks.

Characterizing problem solving tasks

ALL included a total of 4 projects involving 20 tasks in the assessment of problem solving. These resulted in 19 scorable items than ranged from 199 to 394 along the scale and, like the literacy and numeracy tasks, their placement was determined by the patterns of right and wrong responses among adults in participating countries. Rather than release one of the four projects that were used in ALL, we will characterize the hypothesized proficiency scale for analytical problem solving that was tested using pilot data and present an example from the pilot data that

was not used in the main assessment[3]. Similar models have been described within the frameworks of other large-scale assessments of problem-solving competencies such as the project test for Hamburg/Germany (Ebach, Klieme and Hensgen, 2000) and the PISA 2003 assessment of cross-curricular problem solving (OECD, in press).

In ALL, four levels of problem-solving proficiency are postulated:

Level 1

At a very elementary level, concrete, limited tasks can be mastered by applying content-related, practical reasoning. At this level, people will use specific content-related schemata to solve problems.

Level 2

The second level requires at least rudimentary systematical reasoning. Problems at this level are characterized by well-defined, one-dimensional goals; they ask for the evaluation of certain alternatives with regard to transparent, explicitly stated constraints. At this level, people use concrete logical operations.

Level 3

At the third level of problem-solving proficiency, people will be able to use formal operations (e.g., ordering) to integrate multi-dimensional or ill-defined goals, and to cope with non-transparent or multiple dependent constraints.

Level 4

At the final and highest level of competency, people are capable of grasping a system of problem states and possible solutions as a whole. Thus, the consistency of certain criteria, the dependency among multiple sequences of actions and other "meta-features" of a problem situation may be considered systematically. Also, at this stage people are able to explain how and why they arrived at a certain solution. This level of problem-solving competency requires a kind of critical thinking and a certain amount of meta-cognition.

The following example illustrates a concrete realization of a project. For this purpose a project that is not included in the final ALL instrument is introduced and one typical problem-solving task is shown. The project is about "Planning a trip and a family reunion".

In the introductory part of the project, the respondent is given the following summary describing the scenario and overall problem:

> *"Imagine that you live in City A. Your relatives are scattered throughout the country and you would like to organize a family reunion. The reunion will last 1 day. You decide to meet in City B, which is centrally located and accessible to all. Since you and your relatives love hiking, you decide to plan a long hike in a state park close to City B. You have agreed to be responsible for most of the organization."*

The respondent is then given a list of steps he or she needs to work through, in this example the following list:

- *Set the date for the reunion*
- *Consider your relatives' suggestions for the hike*
- *Plan what needs to be done before booking your flight*
- *Answer your relative's questions about traveling by plane*
- *Book your flight*
- *Make sure your ticket is correct*
- *Plan the trip from City B to the airport*

The first task of this project "Set the date for the reunion" is a good example of a typical problem-solving task and is shown here as it would appear in a test booklet.

Example task: Set the date for the reunion

The family reunion should take place sometime in July.

You asked all your relatives to tell you which dates would be suitable. After talking to them, you made a list of your relatives' appointments during the month of July. Your own appointment calendar is lying in front of you. You realize that some of your relatives will have to arrive a day early in order to attend the family reunion and will also only be able to return home on the day after the meeting.

Please look at the list of your relatives' appointments and your own appointment calendar.

List of your relatives' appointments in July 1999

Henry	Karen	Peter	Janet	Anne	Frank
Vacation in City E beginning on July 26; Appointment on July 11	Every day of the week is okay except Thursdays and on July 16	Business appoint-ments on July 2, July 13, and between July 27 and 29	Doesn't have any appoint-ments	Unable to attend reunion on July 5, July 20, or July 24	Has to be away sometime during the 1st full week in July on business, but will find out the exact dates shortly before

Henry, Karen, and Peter could arrive on the same day as the reunion whereas Janet, Anne, and Frank can only arrive on the afternoon before and return home on the day after the reunion.

Example task (cont.)

Your appointment calendar for July 1999

July 1999

	Thurs.	1	Meeting with David
	Fri.	2	
	Sat.	3	
	Sun.	4	
	Mon.	5	
	Tue.	6	
	Wed.	7	
	Thurs.	8	
	Fri.	9	
	Sat.	10	*Hike in City C*
	Sun.	11	
	Mon.	12	
	Tue.	13	
	Wed.	14	
	Thurs.	15	
	Fri.	16	
	Sat	17	
	Sun.	18	
	Mon.	19	
	Tue.	20	
	Wed.	21	
	Thurs	22	
	Fri.	23	
	Sat.	24	
	Sun.	25	
	Mon.	26	
	Tue.	27	
	Wed.	28	Vacation
	Thurs	29	Vacation
	Fri.	30	Vacation
	Sat.	31	

Question 1. Which of the following dates are possible for the family reunion?

Please select all possible dates.

- [a] July 4
- [b] July 7
- [c] July 14
- [d] July 18
- [e] July 25
- [f] July 29

This project illustrates nicely how the action steps logic is actually "translated" into a concrete thematic action flow. The underlying plot – planning a trip and a family reunion – constitutes a very typical everyday-type of action that presumably a large majority of people in different countries will be able to relate to. The action steps themselves and their sequence can deviate from the normative complete action model, as is the case here. The normative model is used as a guideline that is adapted to each specific context. In this case, for example, the task "Consider your relatives' suggestions for the hike" corresponds approximately to the action step "Analyze the situation", the task "Plan what needs to be done before booking your flight" corresponds to the action step "Plan the solution", and "Book your flight" is a typical example for the action step "Execute the plan".

The example task gives a first indication of item structures and formats. The tasks typically start off with a short introduction to the situation, followed by varying types and amounts of information that need to be worked through. In the example task, in order to set the date for the family reunion, the respondent needs to process, compare and integrate the information provided in the list of the relatives' appointments, including the addendum to this list, and their own appointment calendar. Here the information is mostly textual and in the form of tables. The answer format is a multiple-choice format with more than one correct response alternatives, although the number of correct response alternative is not specified.

Conclusion

This paper offers a brief overview of the frameworks that have been used for both developing the tasks used to measure prose and document literacy, numeracy and problem solving in ALL as well as for understanding the meaning of what is being reported with respect to the comparative literacy proficiencies of adults. The frameworks identify a set of variables that have been shown to influence successful performance on a broad array of tasks. Collectively, they provide a means for moving away from interpreting survey results in terms of discrete tasks or a single number, and towards identifying levels of performance sufficiently generalized to have validity across assessments and groups. As concern ceases to center on discrete behaviours or isolated observations and focuses more on providing meaningful interpretations of performance, a higher level of measurement is reached (Messick, 1989).

Endnotes

1. The 80 per cent criterion was drawn from the education literature on mastery learning to reflect a level of performance at which someone is judged to be proficient or competent. Some have argued that this is too high a standard and a response probability of 60 or even 50 per cent should be used. Lowering the criteria to 50 per cent would mean that an adult would be expected to perform tasks at a given level of proficiency with 50 per cent accuracy – hardly a standard we should accept as indicating someone is proficient at something. Would you visit a dentist that fixed the correct tooth 50 per cent of the time? How many employers would hire someone knowing they had a 50/50 chance of performing tasks correctly?
2. Quantitative literacy was defined in IALS as the knowledge and skills needed to apply arithmetic operations either alone or sequentially, using numbers embedded in printed materials.
3. This is the first time problem solving was used in an international survey of adult skills. It is expected that there will be subsequent rounds of ALL and at least some countries will want to measure problem solving using these materials. Therefore, it is important that these four projects be kept confidential for any future use.

References

Almond, R. G., and Mislevy, R. J. (1998). Graphical models and computerized adaptive testing. (TOEFL Tech. Rep. No. 14). Princeton, NJ: Educational Testing Service.

Baker, D., and Street, B. (1994). Literacy and numeracy: Concepts and definitions. In T. Husen and E. A. Postlethwaite (Eds.), Encyclopedia of education. New York: Pergamon Press.

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

Coben, D., O'Donoghue, J., and FitzSimons, G. E. (Eds.)(2000). Perspectives on adults learning mathematics: Theory and practice. London: Kluwer Academic Publishers.

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

Cook-Gumperz, J., and Gumperz, J. (1981). From oral to written culture: The transition to literacy. In M. Whitman (Ed.), Writing: The nature, development and teaching of written communication: Vol. 1. Hillsdale, NJ: Erlbaum.

Crandall, J. (1981, December). Functional literacy of clerical workers: Strategies for minimizing literacy demands and maximizing available information. Paper presented at the annual meeting of the American Association for Applied Linguistics, New York.

Diehl, W. (1980). Functional literacy as a variable construct: An examination of the attitudes, behaviours, and strategies related to occupational literacy. Unpublished doctoral dissertation, Indiana University.

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.

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

Frankenstein, M. (1989). Relearning mathematics: A different third 'R' – Radical maths. London: Free Association Books.

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.

Gal, I. (2000). The numeracy challenge. In I. Gal (Ed.), Adult numeracy development: Theory, research, practice (pp. 1-25). Cresskill, NJ: Hampton Press.

Gal, I. (2002). Adult Statistical literacy: Meanings, components, responsibilities. International Statistical Review, 70(1), 1-25.

Jacob, E. (1982). Literacy on the job: Final report of the ethnographic component of the industrial literacy project. Washington, DC: Center for Applied Linguistics.

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

Heath, S.B. (1980). The functions and uses of literacy. Journal of Communication, 30, 123–133.

Kirsch, I.S., and Guthrie, J.T. (1984a). Adult reading practices for work and leisure. Adult Education Quarterly, 34(4), 213–232.

Kirsch, I.S., and Guthrie, J. T. (1984b). Prose comprehension and text search as a function of reading volume. Reading Research Quarterly, 19, 331–342.

Kirsch, I. (2001). The International Adult Literacy Survey (IALS): Understanding What Was Measured (ETS Research Report RR-01-25). Princeton, NJ: Educational Testing Service.

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

Messick, S. (1989). Validity. In R. Linn (Ed.), Educational measurement (3rd ed.). New York: Macmillan.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. Education Researcher, 32(2), 13-23.

Mikulecky, L. (1982). Job literacy: The relationship between school preparation and workplace actuality. Reading Research Quarterly, 17(3), 400–419.

Miller, P. (1982). Reading demands in a high-technology industry. Journal of Reading, 26(2), 109–115.

Mislevy, R.J. (September, 2000). Leverage points for improving educational assessment. Paper submitted to National Center for Research on Evaluation, Standards, and Student Testing (CRESST) as part of award #R305B60002 from the US Department of Education, Office of Educational Research and Improvement.

Mosenthal, P.B., and Kirsch, I.S. (1998). A new measure for assessing document complexity: The PMOSE/IKIRSCH document readability formula. Journal of Adolescent and Adult Literacy, 41(8), 638–657.

Murray, T.S., Clermont, Y. and Binkley, M. (Eds.) The Adult Literacy and Life Skills Survey: Aspects of Design, Development and Validation. Canada: Statistics Canada, in press.

Murray, T.S., Kirsch, I.S., and Jenkins, L. (1998). Adult Literacy in OECD Countries: Technical report on the First International Adult Literacy Survey. Washington, DC: National Center for Education Statistics.

National Council of Teachers of Mathematics. (2000). Principles and standards for school mathematics. Reston, VA: Author.

Organization for Economic Co-operation and Development. (1992). Adult illiteracy and economic performance. Paris, France: Author.

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

Rychen, D.S. and Salganik, L.H. (Eds.) Key Competencies for a Successful Life and a Well-Functioning Society. Cambridge, MA: Hogrefe and Huber Publishers, 2003.

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

Scribner, S., and Cole, M. (1981). The psychology of literacy. Cambridge, MA: Harvard University Press.

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

Steen, L.A. (2001). Mathematics and democracy: the case for quantitative literacy. USA: National Council on Education and the Disciplines.

Sticht, T.G. (Ed.). (1975). Reading for working: A functional literacy anthology. Alexandria, VA: Human Resources Research Organization.

Sticht, T.G. (1978). Literacy and vocational competency (Occasional Paper 39, National Center for Research in Vocational Education). Columbus, OH: Ohio State University.

Sticht, T.G. (1982, January). Evaluation of the reading potential concept for marginally literate adults. (Final Report FR–ET50–82–2). Alexandria, VA: Human Resources Research Organization.

Szwed, J. (1981). The ethnography of literacy. In M. Whitman (Ed.), Writing: The nature, development, and teaching of written communication: Vol. 1. Hillsdale, NJ: Erlbaum.

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

Venezky, R.L. (1983). The origins of the present-day chasm between adult literacy needs and school literacy instruction. Visible Language, 16, 113–136.

Contributor

Irwin Kirsch, *Educational Testing Service*

Table of Contents

Table of Contents

Table of Contents

Table of Contents

List of Figures

Table of Contents

Table of Contents

Chapter 11

List of Tables

Chapter 2

Chapter 3

Table of Contents

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Chapter 8

Table of Contents

Table of Contents