The International Adult Literacy and Skills Survey, 2003

Public Use Microdata File

User's Manual

Statistics Canada Statistique Canada

Canada

Table of contents

Table of contents, cont'd

List of tables

1.0 Introduction

What is the International Adult Literacy and Skills Survey (IALSS)?

Conducted in 2003, the International Adult Literacy and Skills Survey (IALSS) is the Canadian component of the Adult Literacy and Life Skills program (ALL). The ALL program is a "… large-scale co-operative effort undertaken by governments, national statistical agencies, research institutions and multi-lateral agencies" that provides internationally comparable measures in four domains: prose and document literacy, numeracy and problem solving.[1]

> **What's in a Name?**
>
> ALL...IALSS, why the different names? We get that question all the time. Here's our chance to explain:
>
> There are a number of reasons why the survey name differs in Canada.
>
> For one, sensitivity review uncovered that in Canada, the term "Life-Skills" had specific connotations that were at odds with the types of skills measured in this survey. The concern was that these potential connotations could impact on the response rates and bias the survey.
>
> For another there was a need to register the title of the survey in Canada before the international Program Title (Adult Literacy and Life Skills Survey) had been officially accepted.
>
> The Canadian title "The International Adult Literacy and Skills Survey" also emphasizes the analytical and conceptual link of this new dataset with those of the 1994 IALS.
>
> The added benefit is that we can be much more precise about which database, the IALSS or the ALL, is being used.

The Adult Literacy and Life Skills Program (ALL)

The development and management of the ALL study were co-ordinated by Statistics Canada and the Educational Testing Services (ETS, Princeton, United States) in collaboration with the National Center for Education Statistics (NCES) of the United States Department of Education, the Organization for Economic Co-operation and Development (OECD) and the Institute for Statistics (UIS) of the United Nations Educational, Scientific and Cultural Organization (UNESCO).

The ALL study, undertaken during the first half of 2003, required all participating countries[2] to collect data from a nationally representative sample of at least 3,000 respondents aged 16 to 65 for each language tested – English and French in the case of Canada. The minimum sample requirements for the ALL survey were exceeded in the Canadian IALSS due to contributions by a number of federal departments and provincial governments that allowed for the collection of additional records so as to ensure high reliability in the estimation of data values for sub-national populations. Moreover, unlike the 1994 IALS[3], the 2003 IALSS also benefited from contributions made by territorial governments. As a result, the number of respondents on the full dataset is sufficient to provide accurate estimates for the Yukon,

1 OECD and Statistics Canada (2005), p. 15.

2 Participating countries included Bermuda, Canada, Italy, Norway, Switzerland, the United States and the Mexican state of Nuevo Leon.

3 The International Adult Literacy Survey (IALS) was fielded between 1994 and 1998 in 22 countries. Data for Canada were collected in 1994. Results are reported in three volumes, see OECD and Statistics Canada (1995), OECD and HRDC (1997), and OECD and Statistics Canada (2000).

Northwest Territories and Nunavut. Finally, as with the 1994 IALS, the 2003 IALSS added Canadians over the age of 65 to the sample. Data collection in Canada began in March 2003 and ended in August. By the end of these six months, over 23,000 individuals from across Canada had spent an average of two hours responding to the IALSS and contributing to the value of the survey. Table 1.1 shows the actual and weighted distributions of respondents from across Canada.

Table 1.1 Geographical distribution of IALSS respondents, Canada and jurisdictions, 16 and over, 2003

Province or Territory	Number of respondents	Representative population
Newfoundland and Labrador	1,299	431,647
Prince Edward Island	645	111,274
Nova Scotia	1,272	747,446
New Brunswick	1,466	599,680
Quebec	4,166	5,994,043
Ontario	4,946	9,621,290
Manitoba	2,267	852,805
Saskatchewan	1,234	741,828
Alberta	1,307	2,428,843
British Columbia	1,849	3,313,116
Yukon Territory	1,092	20,738
Northwest Territory	818	26,541
Nunavut	677	12,592
Canada (Total)	**23,038**	**24,901,843**

Please note

The **IALSS** Public Use Microdata File (PUMF) contains the Canadian results of the **ALL** survey, including a number of linguistic and education variables not collected in any of the other participating countries.

On the other hand, the **ALL** Public Use Microdata File contains results from Canada and other participating countries, **but only for respondents aged 16 - 65** and only for a **subset** of the variables present on the **IALSS** dataset. In other words, the International PUMF contains data from other countries, but there are additional variables for Canada and additional respondents (over 65) on the **IALSS** datafile.

In effect, the Canadian component of the **ALL** PUMF is a subset of this Canadian PUMF.

2.0 Survey overview

Every respondent to the IALSS was first given a common questionnaire seeking information about demographic characteristics and variables such as educational attainment, occupation, income, and engagement in adult learning and community activities.

The respondents were then given an internationally validated psychometric[4] instrument designed to measure proficiency in four domains:

- **Prose literacy** – the knowledge and skills needed to understand and use information from texts including editorials, news stories, brochures and instruction manuals.

- **Document literacy** – the knowledge and skills required to locate and use information contained in various formats, including job applications, payroll forms, transportation schedules, maps, tables and charts.

- **Numeracy** – the knowledge and skills required to effectively manage the mathematical demands of diverse situations.

- **Problem solving** – problem solving involves goal-directed thinking and action in situations for which no routine solutions exist. The problem solver has a more or less well defined goal, but it is not immediately obvious 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.

What is Health Literacy

The IALSS Prose, Document and Numeracy domains each contained a number of tasks, such as the oft quoted Aspirin™" bottle, that contained health related tasks and information. By recycling these items and grouping them into a cohesive "domain", it is possible to derive a new set of plausible values that, theoretically as well as empirically, measure the ability to use and understand information from texts, charts and graphs containing Health related information and tasks.

The IALSS dataset includes these measures. They can be used and interpreted in much the same way as the Prose, document and Numeracy domains.

The same prose and document literacy scales were used in both the 1994 IALS and the 2003 IALSS. For both domains, the proficiency scales from the two surveys were linked through the inclusion of a subset of test items originally used in 1994. Thus, for several countries including Canada, the current distributions of prose and document literacy can be compared to those in 1994 to see how these have evolved.

The 2003 IALSS numeracy scale builds on the quantitative literacy domain measured in 1994, providing a broader, more inclusive measure of mathematics skills and conceptual mathematical knowledge. This expanded scale measures more than the ability to perform mathematical operations on numbers embedded in text by including many tasks that require no or little reading.

---

[4] Psychometrics refers to the branch of psychology that deals with the design, administration, and interpretation of quantitative tests for the measurement of variables such as intelligence, aptitude, and personality traits.

Finally, the IALSS carried tasks to assess proficiency in problem solving. This new domain was validated through rigorous testing and displays unique characteristics not found in the other measures. To some extent, it requires the integration of the knowledge and skills measured by the literacy and numeracy domains and their application to new situations. It also implicates basic logical tools needed to provide effective solution strategies to the problems presented in everyday life. These include the ability to order, evaluate and prioritize a series of factors and to discriminate, plan, analyze and reason through a variety of choices in order to arrive at an effective solution to a given problem.

The conceptualization and definitions of the four domains as well as examples of actual test items used in the assessment are presented in the following Chapter. Users requiring additional technical information on the psychometric aspects of the study are referred to Statistics Canada (2004).

3.0 Survey objectives

The IALSS was initiated with two fundamental goals:

1) The first objective was to build on the skills measured in the 1994 IALS by introducing new assessment domains with robust theoretical frameworks and stable item parameters across countries and languages. This goal also involved directly linking the IALS with the IALSS along the two literacy domains in order to allow comparison between prose and document profiles as measured in 1994 and later in 2003.

2) The second objective was to allow for, international, national and sub-national analysis of the correlates and possible antecendents of skills in Canada by collection a large enough bank of information from a sufficiently large number of respondents.

The central element of the survey was the direct assessment of the literacy, numeracy and problem solving skills of respondents using commonplace tasks of varying degree of difficulty drawn from a range of topic and knowledge areas. This information was supported by the collection of background information on respondents. In addition, the background questionnaire included questions on the self-assessment of literacy and numeracy skills of respondents, on the training which the respondent has taken in the year previous to the survey and on the perceived barriers to realizing enhanced literacy or numeracy skill levels.

4.0 Concepts and definitions

4.1 Defining and measuring proficiency: Overview

For IALSS, each proficiency scale starts at zero and increases to a theoretical maximum of 500 points. Scores along the scale denote the points at which a person with a given level of performance has an 80 percent probability of successfully completing a task at that level of difficulty. For instance, a person with an assessed performance at 250 points has an 80 percent probability of correctly answering a task with an estimated difficulty level of 250. The same individual would have an "80 percent plus" probability of correctly answering a simpler task (about 95% for a task with a complexity of 200) and a diminished probability (less than 80%) of successfully completing a more difficult task (about 40% for a task with a complexity of 300).[5]

> **Literacy and illiteracy**
>
> Interestingly, while the probability of a correct response may approach zero as the tasks become more difficult, it can never quite reach it because there is always some chance, however small, that a correct answer will be provided regardless of ability.
>
> Accordingly, the results from the IALSS measure performance along a proficiency continuum. The scales do not measure the absence of a competence, and thus cannot distinguish those who have from those who lack a specific competency.

In an effort to facilitate analysis, these continuous scores have been regrouped into 5 skill levels (only 4 levels were defined for the problem solving scale) with level 1 being the lowest measured level of proficient. The proficiency levels used for IALSS are useful in summarizing the results but also have some limitations. First, the relatively small proportions of respondents who actually reach Level 5 do not always allow for accurate reporting. For this reason, whenever results are presented by proficiency level, Levels 4 and 5 are typically combined. Second, the levels indicate specific sets of abilities and, therefore, the thresholds for the levels are not equidistant. The ranges of scores in each level are therefore not identical. In fact, for all four domains, Level 1 captures almost half of the scale. The thresholds for the problem solving domain are set somewhat differently and Level 1 covers precisely half of the scale. Level 1 includes all basic abilities required to attain higher levels. In other words, the ability to read may lie somewhere in Level 1, but the ability to understand and use what has been read comes in gradations of complexity from Level 1 to Level 5. The upshot of the relatively large ranges of scores in Level 1 on each of the scales is that there are multiple sub-levels of proficiency within this level. The range includes those who can barely read at all as well as those who read poorly or inattentively.[6]

This chapter 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 International Adult Literacy and Skills Survey (IALSS). The importance of developing a framework is thought to be central in construct-based approaches to measurement. Among the

[5] Kirsch, Jungeblut and Campbell (1992), pp. 14-15.

[6] The International Survey of Reading Skills is a follow-up to the 2003 IALSS that will provide more information about respondents at Level 1. Results are expected sometime in 2006.

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 chapter 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.

4.2 Understand what was measured in the International Literacy and Skills Survey

4.2.1 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 IALSS 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 IALSS 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.

4.2.2 Scaling the literacy, numeracy and problem solving tasks in IALSS

The results of the IALSS survey can be 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 IALSS 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 IALSS 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 IALSS 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.

What about Health Literacy?

A fifth scale, closely related to the prose and document domains,called Health Literacy, can also be analysed along this same continuum. However, given that this is a scale created post collection, the framework documentation has yet to formally amalgamate the details from each of the donor domains. As a starting point, we understand by health literacy, the ability to use and understand health related printed information from a variety of sources. As such, whatever drives the difficulty of the prose, document or Numeracy domain items re-used in this new scale contribute to its difficulty as well.

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 IALSS, as in IALS, a high degree of proficiency is defined in terms of a response probability of 80 (RP80). 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.

4.3 Measuring prose and document literacy in IALSS

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 IALSS. 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.

4.3.1 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 IALSS 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.

4.3.2 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.

4.3.3 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.

4.3.4 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.

4.3.5 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 IALSS 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.” Predictably, this item was also used as one of the contributing stimuli for the Health Literacy domain. 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

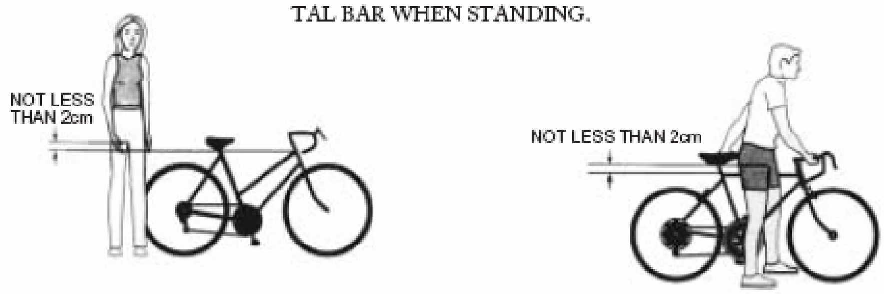
0 67736 11079

* *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.

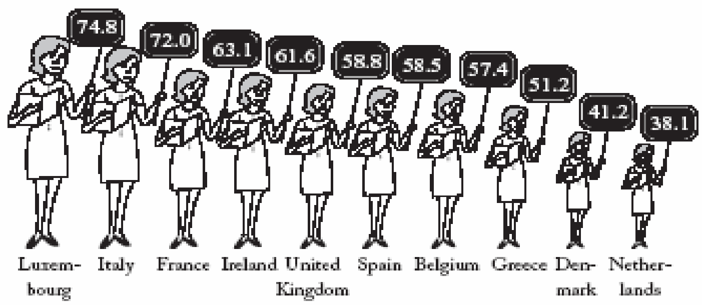
The most difficult task on the prose literacy scale (377) falls in the lower range of Level 5 and required readers to look at an announcement from a personnel department and to "list two ways in which CIEM (an employee support initiative within a company) helps people who lose their jobs because of departmental 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.

4.3.6 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 IALSS. 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

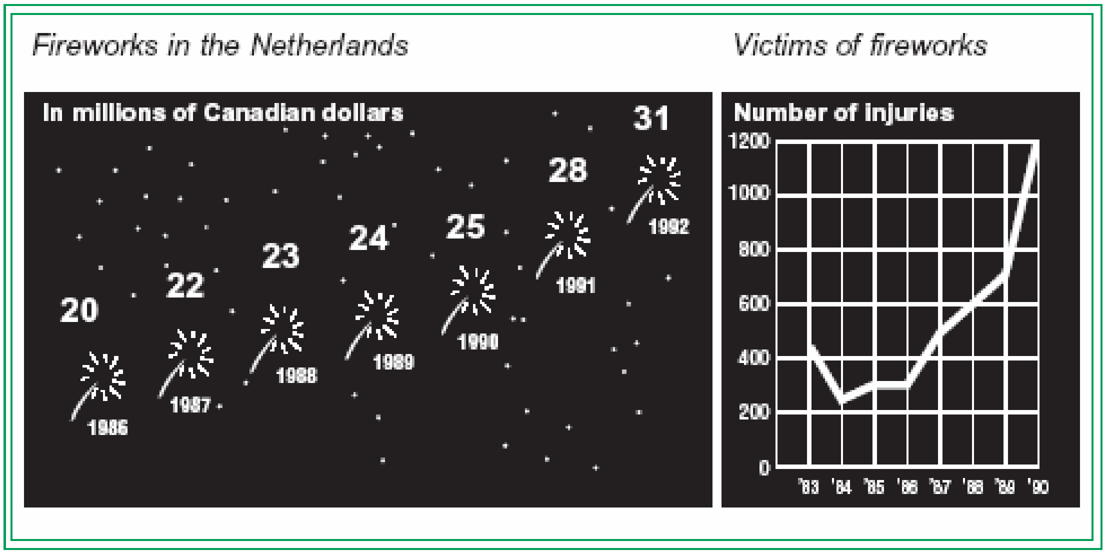
There is a low percentage of women teachers in the Netherlands compared to other countries. In most of the other countries, the majority of teachers are women. However, if we include the figures for inspectors and school principals, the proportion shrinks considerably and women are in a minority everywhere.



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.



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

Better ◉ ◓ ○ ◒ ● Worse

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. 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	2 Price	3 Dimensions (HxWxD cm)	4 Overall score	5 Convenience	6 Performance	7 Sensitivity	8 Selectivity	9 Tone quality	10 Reversible time-setting	11 Dual alarm	Warranty (months)	Advantages	Disadvantages	Comments
Full-featured clock radios														
RCA RP-3690	$50/$40	6x25x18	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	10x26x13	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/—	15x36x13	78	○	◉	◓	◉	◓	—	✔	3	D,G,K,M,O,R,T	b,g	B
Emerson AK2745	39/20	6x28x15	70	○	◒	◒	◒	○	✔	✔	3	G,O	g	K
Soundesign 3753	20/20	6x23x13	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	e	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	15x26x15	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-X260, $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 reset alarm.

Key to Advantages

A – Alarm works despite power failure.
B – Shows actual time plus up to 2 alarm times.
C – Twin alarms suitable 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 – Better than most in tuning ease.
U – Better than most in image rejection.

Key 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 indication alarm is set.
g – Lacks alarm-reset button.
h – Time-setting lacks 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 – 3-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 $3 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

4.4 Measuring numeracy in IALSS

The conception of numeracy developed for IALSS 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 IALSS 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. 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 IALSS 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 IALSS Numeracy Team to serve as the basis for the development of numeracy items for IALSS:

> 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.

4.4.1 Identifying task characteristics

Four key characteristics of numerate behaviour were used to develop and represent the numeracy tasks built for IALSS – 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.

4.4.2 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.

4.4.3 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.

4.4.4 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.

4.4.5 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 IALSS 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.

4.4.6 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.

4.4.7 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.

4.4.8 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.

4.4.9 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.

4.4.10 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.

4.4.11 Characterizing numeracy tasks

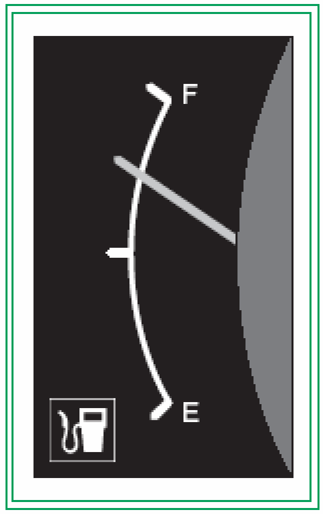
A total of 40 numeracy tasks were selected and used in the IALSS survey. These tasks range along the numeracy scale from 174 to 380 and their placement was determined by how well adults in participating countries responded to each task. Described below are sample tasks that reflect some of the conceptual 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.

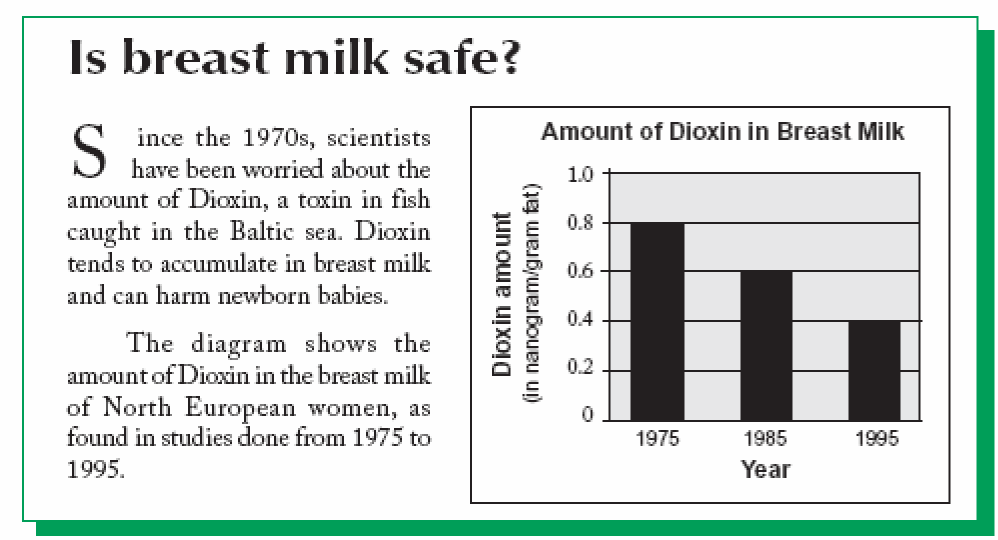
What makes this task more difficult than the previous ones described above is the fact that adults must first estimate the level of gas remaining in the tank, by converting the placement of the needle to a fraction. Then they need to determine how many gallons this represents from the 48 gallon capacity 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). This item also served to help create the Health Literacy Domain.



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.

4.5 Measuring problem solving in IALSS

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 IALSS 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 IALSS 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 IALSS 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.

4.5.1 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 IALSS. 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.

4.5.2 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 IALSS, 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 IALSS, 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 IALSS 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 July 16; Appoint-ment 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 appoin t-ments	Unable to attend reunion on July 5, July 20, or July 24	Has be to away sometime during the 1 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.

4.6 Conclusion

This chapter has offered 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 IALSS 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).

4.6.1 Some analytical considerations

> **Where have all the illiterates gone?**
>
> Like its predecessor, the 2003 IALSS conceptualizes proficiency along a continuum that denotes how well adults use information to function in society and the economy. It bares repeating that the IALSS does not measure the absence of competencies. Rather it measures knowledge and skills in the four domains along a broad range of ability. Consequently, the results cannot be used to classify population groups as either "literate" or "illiterate".

The skill levels presented in the IALSS dataset not only provide a means for exploring the progression of information-processing demands across each of the scales, but also can be used to help explain how the proficiencies individuals demonstrate reflect the likelihood they will respond correctly to the broad range of tasks used in this assessment as well as to any task that has the same characteristics. In practical terms, this means that individuals performing at 250 on each scale are expected to be able to perform the average level 1 and 2 tasks with a high degree of proficiency – i.e. with an average probability of a correct response at 80 per cent or higher. It does not mean that they will not be able to perform tasks in levels 3 or higher. They would be expected to do so some of the time, but not consistently.

Based on the result of the 1994 IALS for the two scales common to the 2003 IALSS, Tables 4.1 and 4.2 display the probability that individuals performing at selected points on the prose or document literacy scales will give a correct response to tasks of varying difficulty. For example, a reader whose prose proficiency is 150 has less than a 50 per cent chance of giving a correct response to the level 1 tasks. Individuals whose proficiency score is 200, in contrast, have about an 80 per cent probability of responding correctly to these tasks.

In terms of task demands, it can be inferred that adults performing at 200 on the prose scale are likely to be able to locate a single piece of information in a brief text when there is no distracting information, or if plausible but incorrect information is present but located away from the correct answer. However, these individuals are likely to encounter far more difficulty with tasks in levels 2 through 5. For example, they would have only a 40 per cent chance of performing the average level 2 task correctly, an 18 per cent chance of success with tasks in level 3, and no more than a 7 per cent chance with tasks in levels 4 and 5.

In contrast, respondents demonstrating a proficiency of 300 on the prose scale have about an 80 per cent chance or higher of succeeding with tasks in levels 1, 2 and 3. This means that they demonstrate success with tasks that require them to make low-level inferences and with those that entail taking some conditional information into account. They can also integrate or compare and contrast information that is easily identified in the text. On the other hand, they are likely to encounter difficulty with tasks where they must make more sophisticated text-based inferences, or where they need to process more abstract types of information. These more difficult tasks may also require them to draw on less familiar or more specialised types of knowledge beyond that given in the text. On average, they have about a 50 per cent probability of performing level 4 tasks correctly; with level 5 tasks, their likelihood of responding correctly decreases to 40 per cent.

Similar kinds of interpretations can be made using the information presented for the document and quantitative literacy scales. For example, someone who is at 200 on the quantitative scale has, on average, a 67 per cent chance of responding correctly to level 1 tasks. His or her likelihood of responding correctly decreases to 47 per cent for level 2 tasks, 21 per cent for level 3 tasks, 6 per cent for level 4 tasks and a mere 2 per cent for level 5 tasks. Similarly, readers with a proficiency of 300 on the quantitative scale would have a probability of 92 per cent or higher of responding correctly to tasks in levels 1 and 2. Their average probability would decrease to 81 per cent for level 3 tasks, 57 per cent for level 4 and 20 per cent for level 5.

Table 4.1 Average probabilities of successful performance, prose scale

	Selected proficiency scores				
Prose level	**150**	**200**	**250**	**300**	**350**
			%		
1	48	81	95	99	100
2	14	40	76	94	99
3	6	18	46	78	93
4	2	7	21	50	80
5*	2	6	18	40	68

* Based on one task

Source: Adult Literacy Survey (1994).

Table 4.2 Average probabilities of successful performance, document scale

	Selected proficiency scores				
Document Level	**150**	**200**	**250**	**300**	**350**
			%		
1	40	72	94	99	100
2	20	51	82	95	99
3	7	21	50	80	94
4	4	13	34	64	85
5*	<1	1	3	13	41

* Based on one task

Source: Adult Literacy Survey (1994).

Proficiency in each domain is measured on a continuous scale. Each scale starts at zero and increases to a theoretical maximum of 500 points (with four decimal places of precision). Scores along the scale denote the points at which a person with a given level of performance has an 80 percent probability of successfully completing a task at that level of difficulty.

From an analytical standpoint, these continuous measures can be useful in creating summary statistics that describe the competencies of populations such as their overall average score. Populations with similar average scores, however, may have quite different numbers of low or high performing adults. Thus, one can also look at how the scores are distributed within populations by using percentile scores. Percentile scores are the scores below which a specified percentage of adults are found. Thus, for example, the 5th percentile score is the one below which we find 5% of adults in a particular population. Differences in percentile scores tell us something about the degree of equality in proficiency across populations. Users should refer to chapter 5 for more detailed information on the correct use of the plausible values in their analysis.

The IALSS scores have also been grouped into proficiency levels representing tasks of increasing difficulty. For the prose and document literacy domains as well as the numeracy domain, experts have defined five broad levels of difficulty, each corresponding to a similar (though not equidistant) range of scores. For the problem solving domain, experts have defined four broad levels of difficulty. In each domain, Level 1 denotes the group with the lowest proficiency and Level 4 for the problem solving and 5 for the other domains contains the highest performers.

It is important, for analytical as well as operational reasons, to define a “desired level” of competence for coping with the increasing skill demands of the emerging knowledge and information economy. Level 3 performance is generally chosen as a benchmark because in developed countries, performance above Level 2 is generally associated with a number of positive outcomes. These include increased civic participation, increased economic success and independence, and enhanced opportunities for lifelong learning and personal literacy (Kirsch, I., et al., 1993; Murray, T.S. et al., 1997; Tuijnman, A., 2001). Whereas individuals at proficiency Levels 1 and 2 typically have not yet mastered the minimum foundation of literacy needed to attain higher levels of performance (Strucker, J., Yamamoto, K. 2005)

Secondary analysis of the 1994 IALS data has yielded consistent evidence that the performance difference between Level 2 and Level 3 on the prose, document and quantitative literacy scales is substantive and corresponds to a significant difference in measurable benefits accruing to citizens in OECD countries (OECD and HRDC, 1997). Results of preliminary analysis of the IALSS data, including the new numeracy scale, are consistent with this finding. For this reason, it is sometimes useful to anchor the scales at the cut point between Levels 2 and 3, thus highlighting the distributions above and below this threshold for the prose, document and numeracy domains. In contrast, interpretation of the problem solving domain is more complex and no single “desirable” threshold has yet been set, in which case, a cutpoint at level 1 may be more appropriate until a more precise threshold can be found.

Chapter 8 will offer further tools and techniques for the appropriate use of the Plausible Values and Replicate Weights required to produce accurate estimates of the standards errors associated with each point estimate.

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.

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

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.

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.

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

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

5.0 Survey methodology

As a country participating in the first round of data collection for the international Adult Literacy and Life Skills (ALL) survey, the Canadian component, under the heading of the International Adult Literacy and Skills Survey (IALSS), was implemented according to the standards provided in the document '*Standards and Guidelines for the Design and Implementation of the Adult Literacy and Life Skills Survey*'. These standards establish the minimum design and implementation requirements covering the complete range from survey planning to survey documentation.

5.1 Assessment design

As discussed in the previous chapter, the elements of functional literacy and other skills in IALSS are evaluated through psychometric measures of proficiency in the skill domains of prose literacy, document literacy, numeracy, and problem solving. Every question, or set of related questions, is based on an item. The set of all items are organized into smaller sets of tasks, or blocks: four 30-minute blocks of literacy items (i.e., prose and document literacy), two 30-minute blocks of numeracy items, and two 30-minute blocks of problem solving items. The blocks are combined in pairs using a Balanced Incomplete Block (BIB) assessment design to arrive at 28 booklets.

The booklets were distributed amongst the sample according to the design for the entire Canadian sample, over and beyond the minimum requirement of respondents from each language tested. As each booklet can take upwards of an hour to administer, each respondent was asked to complete only one; no one was required to take the entire set of tasks. The method of spreading the blocks across booklets substantially reduced the burden on respondents. The data collection activity was also closely monitored in order to obtain approximately the same number of complete cases for each task booklet, except for four task booklets containing either only numeracy items or only problem solving items: these booklets required a larger number of complete cases.

5.2 Population and coverage

The target population is comprised of all Canadian residents who were 16 years of age or older at the time of data collection, excluding long-term institutional residents, families of members of the armed forces on military bases, and individuals living on Indian Reserves.

Residents of sparsely populated regions were also excluded from the survey population for operational reasons. Even when combined with the exclusions listed above, this represented no more than 2 per cent of the total population, well within the international 5 per cent minimum under-coverage requirement. It is estimated that the coverage for the survey was 98.5 per cent nationally, with provincial coverage ranging from 95 per cent to nearly 100 per cent. In the northern territories, reduced levels of coverage (70-90%) were obtained because only the communities covered in the national Labour Force Survey were included. Table 5.1 provides the estimated coverage rate by province and territory.

Table 5.1 Estimated coverage rate by jurisdiction, IALSS 2003

	Estimated coverage rate (%)
Newfoundland and Labrador	98.1
Prince Edward Island	99.7
Nova Scotia	99.3
New Brunswick	98,8
Québec	98,9
Ontario	99,3
Manitoba	95,3
Saskatchewan	95,3
Alberta	98,2
British Columbia	97.1
Yukon	90.0
Northwest Territories	86.0
Nunavut	70.0
Canada	**98.5**

Source: International Adult Literacy and Skills Survey, 2003.

Given the reduced coverage in the North, we did not inflate the weights (through calibration) to territorial demographic totals. Instead we calibrated to control totals covering the communities included in our sample. The samples in the North therefore represent only the specific communities covered by the same. Here are the communities covered:

Yukon: Whitehorse, Whitehorse Unorganized (i.e. rural Whitehorse), Mt. Lorne, Ibex Valley, Dawson, Watson Lake, Haines Junction, Carmacks, Mayo, Ross River and Carcross.

Northwest Territories: Yellowknife, Hay River, Inuvik, Norman Wells, Fort Smith, Fort Simpson, Rae-Edzo, Fort Providence, Fort Liard, Fort Resolution, Tuktoyaktuk, Fort McPherson and Aklavik.

Nunavut: Iqaluit, Rankin Inlet, Cambridge Bay, Kugluktuk, Arviat, Baker Lake, Pangnirtung, Cape Dorset, Igloolik and Pond Inlet.

Canada was the only country that opted to include adults over the age of 65 in its target population; a liberty that was available as the sample design already satisfied the minimum suggested international sample size requirement for those aged 16 to 65 years.

5.3 Survey frame

The most recent Census of Population and Housing, with a reference date of May 15th 2001, was chosen as the frame for the survey. This already existing frame offered the ability to use reported household-level characteristics to identify dwellings with greater probability of containing an individual belonging to specific target subpopulations of interest. This auxiliary information greatly assisted the efficiency of the sample design. Specifically, the survey frame consisted of households enumerated by the Census long-form (20%) sample.

The survey's national base sample, provincial top-up samples to the base, and supplementary samples related to age could have been selected from short-form households, but the long form data was required to identify the remainder of the special subpopulations. In the case of minority language samples, the quality of the long form responses is judged to be superior to that of the short form. The presence of questions on the knowledge of, and the use of languages, in addition to the mother tongue (language first learned and still understood) provide respondents with more opportunities to properly characterize their linguistic profile.

5.4 Sample design

A stratified multi-stage probability sample design was used to select the sample from the Census Frame. The sample was designed to yield separate samples for the two official languages, English and French. In addition, the sample size was increased to produce estimates for a number of population subgroups. Provincial ministries and other organizations sponsored supplementary samples to increase the base or to target specific subpopulations such as youth (ages 16 to 24 in Québec and 16 to 29 in British Columbia), adults aged 25 to 64 in Québec, linguistic minorities (English in Québec and French elsewhere), recent and established immigrants, urban aboriginal peoples, and residents of the northern territories.

In each of the 10 provinces the Census Frame was further stratified into an urban stratum and a rural stratum. The urban stratum was restricted to urban centers of a particular size, as determined from the previous census. The remainder of the survey frame was delineated into primary sampling units (PSUs) by Statistics Canada's Generalised Area Delineation System (GArDS). The PSUs were created to contain a sufficient population in terms of the number of dwellings within a limited area of reasonable compactness. In addition, a general indication of the education level of the population from the 1996 Census was incorporated to create PSUs that reflected the educational distribution of their province.

A second, implicit, stratification was used in the systematic selection of households for each sample. The highest level of education for each adult in the household, as recorded in the Census frame, was used to determine a representation of the dominant class from four broad levels: 1) less than high school, 2) high school graduate or some post-secondary education, 3) college graduate, and 4) university graduate. Formal educational attainment is not the only, but is the main, determinant of performance in evaluations of literacy (OECD and Statistics Canada, 2000). Ordering the households by education within geographic regions before sample selection increased the ability to represent a range of educational backgrounds.

The sample was allocated between strata under a Neyman allocation, incorporating a conservative design effect of 2 for the rural stratum and 1.5 for the urban stratum. After allocation, it became apparent that several PSUs in the rural strata were sufficiently important that they were effectively being sampled with certainty. These PSUs were converted to a new pseudo-urban stratum, to be treated similar to the urban stratum in terms of sample selection.

As a final step before sample selection, the negotiated sample sizes were inflated to account for an international target minimum response rate of 70 per cent and for mobility in terms of the characteristics of interest for each subpopulation covered by a supplementary sample. A blended rate was calculated using reported 1-year and 5-year mobility variables from the Census as proxy variables, and applied to the time lag between the Census and the start of collection in March of 2003. These rates were adjusted downward in each stratum to reflect the expected replacement of movers by others with the same target characteristics for each supplementary sample.

Within the urban stratum, two stages of sampling were used. In the first stage, households were selected systematically with probability proportional to size. The size measure was constructed in terms of the number of adults in a household, using a maximum cap at four for the base sample and at three for supplementary samples. During the second stage, a simple random sample algorithm was used by the CAPI application to select an individual from the demographic roster of eligible household adults. Three stages were used to select the sample in the rural stratum. In the first stage, primary sampling units (PSUs) were selected with probability proportional to population size as measured by the total number of adults for each sample's survey population in the 2001 Census. The second and third stages for the rural stratum repeated the same methodology employed in the two-stage selection for the urban stratum.

5.5 Required precision

Three basic levels of sample sizes were proposed for the survey, a minimum, an average, and a maximum strategy depending on the precision required from each estimate. The minimum strategy corresponded to the ability to estimate a proportion (such as the proportion of the population at a given level of proficiency for a given domain) as low as 15% with a coefficient of variation of at most 16.5%, i.e., of acceptable quality. In the 1994 IALS, the smallest proportions of Canadian adults in any published skill level were found in the lowest level of proficiency; 16.6% were classed in Level 1 for Prose Literacy and a similar amount, 16.9% for Quantitative Literacy (numeracy). The Problem Solving domain was not part of IALS and so a good national estimate of this proportion will only be known once the estimates from the IALSS are compiled. The average strategy is very similar to the minimum, except that the ability to estimate a proportion as small as 10% was sought. Finally, the maximum strategy corresponded to the ability to produce estimates for proportions as small as 25% *within* a literacy level, provided that level represents at least 15% of the total population.

5.6 Sample size

Each province had a base sample that covered the general population. Additionally, provincial ministries and other organisations sponsored supplementary samples to increase the base or to target specific subpopulations. Table 5.2 shows the expected number of respondents in each sample: the base, youth (ages 16-24 in Québec and 16-29 in British Columbia), adults aged 25-64 in Québec, linguistic minorities (English in Québec and French elsewhere), recent and established immigrants, urban aboriginals, and residents (specifically Inuit and non-Inuit for Nunavut) of the territories.

Table 5.2a Expected distribution of responses prior to fielding, IALSS 2003

Jurisdiction	Base sample	Youth*	Adult**	Language	Immigrant	Aboriginal populations	Non- aboriginal populations***	Total
Newfoundland and Labrador	1,350	…	…	…	…	…	…	1,350
Prince Edward Island	650	…	…	…	…	…	…	650
Nova Scotia	1,350	…	…	…	…	…	…	1,350
New Brunswick	650	…	…	760	…	…	…	1,410
Québec	1,110	815	1,885	570	270	…	…	4,650
Ontario	1,690	…	…	3,000	1,060	…	…	5,750
Manitoba	1,350	…	…	450	…	700	…	2,500
Saskatchewan	650	…	…	…	…	700	…	1,350
Alberta	1,350	…	…	…	70	…	…	1,420
British Columbia	1,350	490	…	…	280	…	…	2,120
Yukon	…	…	…	…	…	700	700	1,400
Northwest Territories	…	…	…	…	…	450	450	900
Nunavut	…	…	…	…	…	700	180	880
Canada	**11,500**	**1,305**	**1,885**	**4,780**	**1,680**	**3,240**	**1,340**	**25,730**

Table 5.2b Final distribution of respondents, IALSS 2003

Jurisdiction	Base sample	Youth*	Adult**	Languag e	Immigrant	Aboriginal populations	Non- aboriginal populations***	Total
Newfoundland and Labrador	1,299	…	…	…	…	…	…	1,350
Prince Edward Island	645	…	…	…	…	…	…	650
Nova Scotia	1,272	…	…	…	…	…	…	1,350
New Brunswick	636	…	…	830	…	…	…	1,410
Québec	1,002	693	1,737	465	269	…	…	4,650
Ontario	1,502	…	…	2,318	1,126	…	…	5,750
Manitoba	1,219	…	…	386	…	662	…	2,500
Saskatchewan	559	…	…	…	…	675	…	1,350
Alberta	1,234	…	…	…	73	…	…	1,420
British Columbia	1,190	345	…	…	314	…	…	2,120
Yukon	…	…	…	…	…	615	477	1,400
Northwest Territories	…	…	…	…	…	408	410	900
Nunavut	…	…	…	…	…	570	107	880
Canada	**10,558**	**1,038**	**1,737**	**3,990**	**1,782**	**2,930**	**994**	**23,038**

* Youth = 16 to 24 in Quebec, 16 to 29 in British Columbia
** Adults are defined as being 25 to 64
*** Non-aboriginal population in Nunavut is defined as anyone who is Non-Inuit
Source: International Adult Literacy and Skills Survey, 2003

After adjusting for non-response and the anticipated mobility of the target sub-populations, an overall sample size of over 40,000 was achieved. The samples were selected sequentially, one after another, starting with the base sample. After the selection of each sample, chosen households were removed from the frame before the next selections, thereby making the samples dependent. The sequential selection of multiple samples in a province can be viewed as multiple phase sampling.

A respondent's data is considered complete for the purposes of the scaling of psychometric assessment data provided that at least the Background Questionnaire variables for age, gender and education have been completed.

5.7 Data collection

The ALL survey design combined educational testing techniques with those of household survey research to measure literacy and provide the information necessary to make these measures meaningful. The main task booklets were the last of a series of collection instruments to be applied. Initially, respondents were asked to complete a survey entry component, or screener, which constructed a roster for each sampled dwelling. This screener collected enough demographic data to identify target sub-populations for the survey and to permit the random selection of one member from each dwelling. The background questionnaire was then asked of the selected respondent, encompassing several modules of information required to relate the tested skills to individuals' economic and social situations. The respondents were asked a series of questions on educational attainment, literacy practices at home and at work, labour force information, information communications technology uses, adult education participation and literacy self-assessment. As a result, the background questionnaire required a median time of about 35 minutes to administer.

Once the background questionnaire had been completed, the interviewer presented a short core task booklet of six relatively simple tasks (Core task booklet). Respondents who passed the Core tasks were given a more difficult main task booklet, with a much larger variety of tasks involving about 45 items. No time limit was imposed on respondents, and they were urged to try each item in their booklets. Respondents were given a maximum leeway to demonstrate their skill levels, even if their measured skills were minimal. All respondents were to attempt the Core task, and then if indicated, the main task booklet (median completed time of 58 minutes) immediately after completing the background questionnaire in order to control the impact of fatigue on the assessment tools.

The core and main task booklets were paper and pencil assessments; however the screener, background questionnaire, and even the administration of the core and the main task booklets were handled in a computer-assisted personal interview (CAPI) environment. As a benefit of an extensive match of the census frame to the central Address Register, telephone numbers were available for approximately 74 per cent of the survey file. In such a case, interviewers were permitted to make an initial contact by telephone to complete the screener and to then schedule an appointment for a personal interview with the selected respondent.

Data collection began in 2003 with the planning of interviewer assignments by the regional offices coordinating the collection activities. The first contacts with respondents were initiated in March 2003 across the country and the last interviews were completed in August, with all survey-related materials being returned to head office by September of 2003.

To ensure high quality data, the international Survey Administration Guidelines were followed and supplemented by adherence to Statistics Canada's own internal policies and procedures. The interviews were conducted in the respondent's home in a neutral, non-pressured manner. Interviewer training and supervision were provided, emphasizing the importance of precautions against non-response bias. Interviewers were specifically instructed to return several times to non-respondent households in order to obtain as many responses as possible. Extensive effort was expended to ensure that the address information provided to interviewers was as complete as possible, in order to reduce potential household identification problems. Finally, the interviewers' work was supervised by using frequent quality checks at the beginning of data collection, fewer quality checks throughout collection and having help available to interviewers during the data collection period. In total, Canada used 317 interviewers with an average assignment size of 62 respondents.

As a condition of participation in the international study, it was required to capture and process files using procedures that ensured logical consistency and acceptable levels of data capture error. Specifically, complete verification of the captured scores (i.e., enter each record twice) was done in order to minimize error rates. Because the process of accurately capturing the task scores was essential to high data quality, 100 per cent keystroke verification was required.

Industry, occupation, and education variables were required to be coded using standard schemes such as the International Standard Industrial Classification (ISIC), the International Standard Classification of Occupations (ISCO) and the International Standard Classification for Education (ISCED). Coding schemes were provided for all open-ended items, as were specific instructions about coding of such items.

5.8 Scoring of tasks

Persons charged with scoring received intense training in scoring responses to the open-ended items using the ALL scoring manual. As well, they were provided a tool for capturing closed format questions. To aid in maintaining scoring accuracy and comparability between countries, the ALL survey introduced the use of an electronic bulletin board, where countries could post their scoring questions and receive scoring decisions from the domain experts. This information could be seen by all participating countries, and they could then adjust their scoring. To further ensure quality, monitoring of the scoring was done in two ways.

First, at least 20 per cent of the tasks had to be re-scored. Guidelines for intra-country rescoring involved rescoring a larger portion of booklets at the beginning of the scoring process to identify and rectify as many scoring problems as possible. In a second phase, a smaller portion of the next third of the scoring booklets was selected; the last phase was viewed as a quality monitoring measure, which involved rescoring a smaller portion of booklets regularly to the end of the re-scoring activities. The two sets of scores needed to match with at least 95 percent accuracy before the next step of processing could begin. In fact, most of the intra-country scoring reliabilities were above 95 per cent. Where errors occurred, a country was required to go back to the booklets and rescore all the questions with problems and all the tasks that belonged to a problem scorer.

Table 6.1 in the next chapter the high level of intra-country score agreement that was achieved.

Second, an international re-score was performed (inter-country rescore). Each country had 10 per cent of its sample re-scored by scorers in another country. For example, a sample of task booklets from the United States was re-scored by the persons who had scored Canadian English booklets, and vice-versa. The main goal of the re-score was to verify that no country scored consistently differently from another. Inter-country score reliabilities were calculated by Statistics Canada and the results were evaluated by the Educational Testing Service based in Princeton. Again, strict accuracy was demanded: a 90 per cent correspondence was required before the scores were deemed acceptable. Any problems detected had to be re-scored.

The details of the intra and inter country rescores will be revisited in the next chapter.

5.9 Survey response and weighting

The Canadian IALSS sample has a very complex design, involving stratification, multiple phases, multiple stages, systematic sampling, probability proportional to size sampling, and several overlapping samples. Furthermore, there is a need to compensate for the non-response that occurred at varying levels. Therefore, the estimation of population parameters and the associated standard errors is dependent on the survey weights. Two types of weights were calculated: population weights that are required for the production of population estimates, and jackknife replicate weights that are used to derive the corresponding standard errors.

5.9.1 Population weights

The population weights were derived in four steps: 1) calculation of the design weights, 2) weighting adjustments for non-response, 3) integration of the weights from the different samples, and 4) calibration.

The design weights were defined as the inverse of the probabilities of selection. The overall probability of selection of a sample unit was the product of its probabilities of selection at each phase and stage of selection. The sequential selection of multiple samples in a province was taken into account by factoring in the probability that a unit selected in a given sample was not selected in any of the samples already selected.

The weighting adjustments for non-response were calculated by first categorizing the sample units either as respondents, out-of-scope households, non-respondent households (those without data from the screener), and non-respondent individuals (screener completed, but no data for the selected respondent). The CHAID algorithm in Knowledge-Seeker software was used successively to form weighting classes (response homogeneous groups) to adjust for non-respondent households and non-responding persons in two separate stages for each province and sample type. Afterward, the design weights of the respondents were adjusted by the factors calculated from each step in order to represent all individuals.

With the overlap in coverage from the various samples, it was necessary to integrate the weights to be able to produce estimates using all units from all samples. The situation is comparable to a multiple frame situation, except that here the samples are dependent. The weights were integrated using Hartley's method for multiple frames: the entire sample was partitioned according to the sub-populations targeted in the supplementary samples, and the weights were adjusted by coefficients proportional to the realized sample sizes of the various samples within the partition.

Finally, the weights were calibrated separately in each province or territory using the benchmark variables given in Table 5.3. Attempts to include household size and education variables proved unsatisfactory and were abandoned. Variables that were used had been validated through matches of the collected survey data with available frame information. Small amounts of missing data for the calibration variables were imputed. Census counts for all calibration variables at the enumeration area level were inflated according to the growth measured between provincial age and gender totals from the Census and the corresponding official demographic counts as of June 21, 2003. This reference date represented an approximation of the midpoint of collection both in terms of calendar days, and in terms of completed response.

Table 5.3 Benchmark variables by province or territory

Jurisdiction	Calibration variables
Newfoundland and Labrador	Age groupe x Gender, Stratum x Gender, CMA/CA (St. John's)
Prince Edward Island	Age groupe x Gender
Nova Scotia	Age groupe x Gender, Stratum x Gender, CMA/CA (Halifax)
New Brunswick	Age groupe x Gender, Stratum x Gender, Francophone x Gender
Québec	Age groupe x Gender, Stratum x Gender, CMA/CA (Montréal, Québec), Anglophone x Gender, Immigrant x Gender
Ontario	Age groupe x Gender, Stratum x Gender, CMA/CA (Toronto, Ottawa, group of 6 CMAs), Francophone x Gender, Immigrant x Gender
Manitoba	Age groupe x Gender, Stratum x Gender, CMA/CA (Winnipeg), Francophone x Gender, Urban Aboriginal x Gender
Saskatchewan	Age groupe x Gender, Stratum x Gender, CMA/CA (Regina, Saskatoon), Urban Aboriginal x Gender
Alberta	Age groupe x Gender, Stratum x Gender, CMA/CA (Calgary, Edmonton), Immigrant x Gender
British Columbia	Age groupe x Gender, Stratum x Gender, CMA/CA (Vancouver), Immigrant x Gender
Yukon	Age groupe x Gender,CMA/CA (Whitehorse), Aboriginal x Gender
Northwest Territories	Age groupe x Gender,CMA/CA (Yellowknife), Aboriginal x Gender
Nunavut	Age groupe x Gender, Inuits x Gender

Note: Age group was defined as 16 to 25, 26 to 35, 36 to 45, 46 to 55, 56 to 65, 66 and over, except that the younger age groups were defined as 16 to 24, 25 to 35 for Québec, and 16 to 29, 30 to 45 for British Columbia.

Source: International Adult Literacy and Skills Survey, 2003.

Table 5.4 summarizes the sample sizes and response rates for each province:

Table 5.4 Achieved sample and response rates by province

Region	Sample size and response rate summary				
	Target population	Initial sample size	Out-of-scope cases[1]	Number of respondents	Response rate[3]
Newfoundland and Labrador	431,646	2,001	98	1,299	68.3
Prince Edward Island	111,274	929	48	645	73.2
Nova Scotia	747,447	1,928	103	1,272	69.7
New Brunswick	599,679	2,126	181	1,466	75.4
Québec	5,994,042	7,327	939	4,166	65.2
Ontario	9,621,290	9,600	1,613	4,946	61.9
Manitoba	852,805	4,186	767	2,267	66.3
Saskatchewan	741,829	2,542	640	1,234	64.9
Alberta	2,428,842	2,067	130	1,307	67.5
British Columbia	3,313,115	3,291	429	1,849	64.6
Yukon	20,739	2,000	249	1,092	62.4
Northwest Territories	26,541	1,286	110	818	69.6
Nunavut	12,592	1,257	119	677	59.5
Canada	**24,901,841**	**40,540**	**5,426**	**23,038**	**65.6**

1. Out-of-scope cases are those that were coded as residents not eligible, unable to locate the dwelling, dwelling under construction, vacant or seasonal dwelling, or duplicate cases.
2. A respondent's data is considered complete for the purposes of the scaling psychometric assessment data provided that at least the Background Questionnaire variables for age, gender and education have been completed.
3. The response rate is calculated as number of respondents divided by the initial sample size minus the out-of-scope cases.
Source: International Adult Literacy and Skills Survey, 2003.

5.9.2 Jackknife weights

To simplify variance estimation, it was assumed that the various samples were selected independently. It is believed that this assumption causes a slight overestimation of the variance. The assumption allowed the use of jackknife variance estimation as proposed by Lohr and Rao (1997) for multiple frames. They propose treating the samples from the different frames as samples from different strata, and apply the jackknife as for stratified sampling. To meet international standards, 30 jackknife replicate weights that cut across strata are included on the survey's micro-data file.

These thirty jackknife replicate weights were developed for use in determining the standard errors of the survey estimates. The 30 replicates cut across strata – 30 replicates were created within each province by sample type and by stratum.

In the urban strata, households were selected systematically in the first stage of sample selection. The replicates were formed by sorting the households in the order that was used for the systematic sample selection, and by assigning replicate numbers sequentially from 1 to 30 to the households, restarting back to 1 after reaching 30.

In the rural strata, PSUs were selected in the first stage of selection. Since fewer than 30 PSUs were selected in all strata, the PSUs were split to form the replicates. As much as possible, the PSUs were split into an equal number of replicates. If this was not possible, the PSUs with a larger number of respondents were split into more replicates. For example, if 11 PSUs were selected for a particular province and sample, then the 8 PSUs with the greatest number of respondents would each be split into 3 replicates, and the remaining 3 PSUs would each be split into 2 replicates, to obtain a total of 30 replicates.

The replicates were formed using the initial IALSS sample of over 40,000 units. Initial jackknife weights were calculated, based on the design weights. The entire weighting process was repeated for each of the 30 jackknife weights, including non-response weighting adjustments, integration of the weights, and calibration.

References

Lohr, S. and Rao, J.N.K. (1997). Jackknife Variance Estimation in Multiple Frame Surveys. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, pp. 552-557.

Hartley, H.O. (1962). Multiple Frame Surveys. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, pp. 203-206.

Organisation for Economic Co-operation and Development, OECD (2000). *Literacy in the Information Age: Final Report on the International Adult Literacy Survey*. Ottawa: Statistics Canada.

OECD and Statistics Canada (2000), *Literacy in the information age: Final results of the International Adult Literacy Survey.* Paris and Ottawa: Authors.

Särndal, C.-E., Swenson, B., and Wretman, J.H. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Special Surveys Division (1999). *International Adult Literacy and Skills Survey: National Planning Report*. Ottawa: Statistics Canada.

6.0 Survey procedures and data processing

6.1 Introduction

In many respects, the IALSS procedures were guided by the international guidelines for the administration of the ALL survey. Standard instruments, sampling, collection and processing methodology (including standardized code for Occupation, Industry and Education) are each important components in making the IALSS part of an Internationally comparative study program. The following section outlines these procedures and details any deviations from the protocol in Canada. The section will also look at some of the details regarding the post-collection data processing leading up to the creation of the Public Use Microdata File.

The IALSS gathered descriptive and proficiency information from sampled respondents through a background questionnaire and a series of assessment booklets containing prose, document, numeracy and problem solving tasks. Survey respondents spent approximately 30 minutes answering a common set of background questions concerning their demographic characteristics, educational experiences, labor market experiences, and literacy related activities. Responses to these background questions make it possible to summarize the survey results using an array of descriptive variables, and also increase the accuracy of the proficiency estimates for various subpopulations. Background information was collected by trained interviewers.

After answering the background questions, the remainder of respondents' time was spent completing a booklet of designed to measure their prose, document, numeracy and problem solving skills. Most of these tasks were open-ended; that is, they required respondents to provide a written answer.

To achieve good content coverage of each of four skill domains, the number of tasks in the assessment had to be quite large. Yet, the time burden for each respondent also needed to be kept within an acceptable range. To accommodate these two conflicting requirements—in other words, to reduce respondents' time burden without sacrificing good representation of the content domain—each respondent was administered only a fraction of the pool of tasks, using a variant of matrix sampling.

6.2 Model procedures manuals and instruments

Each IALSS country was given a set of administration manuals and survey instruments to use as a model. Countries were permitted to adapt these models to their own national data collection systems, but they wer required to retain a number of key features. First, respondents were to complete the core and main test booklets alone, in their homes, without help from another person or from a calculator. Second, respondents were not to be given monetary incentives for participating. Third, despite the prohibition on monetary incentives, interviewers were provided with procedures to maximize the number of complete background questionnaires, and were to use a common set of coding specifications to deal with non-response. This last requirement is critical. Because non-completion of the core and main taks booklets is correlated with ability, background information about non-respondents is needed in order to impute cognitive data for these persons.

6.2.1 Background questions

The model background questionnaires given to each country contained two sets of questions: mandatory questions, which all countries were required to include; and optional questions, which were recommended but not required. Countries were not required to field literal translations of the mandatory questions, but were asked to respect the conceptual intent of each question in adapting it for use. Countries were permitted to add questions to their background questionnaires if the additional burden on respondents would not reduce response rates.

In Canada, the IALSS fielded all of the mandatory and optional questions, as well as a group of idiosyncratic questions of national interest mostly related to mother tongue and immigrant or aboriginal status.

Section A contains variables related to the respondents educational background
Section B contains variables related to linguistic information
Section C contains variables related to parental information
Section D contains variables related to the respondents work experience
Section E contains variables related to the use of different skills at work
Section F contains variables related to continued adult education
Section G contains variables on personal skill practices and health
Section H contains variables related to internat access and usage
Section K contains variables related to personal and household income

Where the answers to these questions do not compromise the confidentiality of our respondents, the IALSS PUMF includes as much of the collected details as possible. Chapter 8 will examine the issues surrounding confidentiality in more details.

6.2.2 Tasks items

Like the IALS before it, the ALL study is based on the premise that the difficulty of various literacy tasks is determined by certain factors, which are stable across language and culture. Accordingly, all of the ALL countries were given graphic files containing the pool of psychometric items and were instructed to modify each item by translating and adapting the English text to their own language without altering the graphic representation or task characteristics. In many cases, the original item was itself translated into this English model providing everyone with the same starting point. This consistency in the base materials minimized the effects of translation and adaptation errors.

Certain rules governed the item modification process. For instance, some items required respondents to perform a task that was facilitated by the use of keywords. In some cases, the keywords were identical in the question and the body of the item; in others, the keyword was similar but not exactly the same; and in still other cases, the keyword was a synonym of the word used in the body of the item. In another case, respondents were asked to choose among

multiple keywords in the body of the item, only one of which was correct. Countries were required to preserve these conceptual associations during the translation process.
Particular conventions used in the items—for example, currency units, date formats, and decimal delimiters—were adapted as appropriate for each country.

To ensure that the adaptation process did not compromise the psychometric integrity of the items, each country's test booklets were carefully reviewed for errors of adaptation.

6.2.3 Standardized non-response coding

It was crucial that the ALL countries managed non-respondent cases in a uniform manner so as to limit the level of non-response bias in the resulting survey estimates.

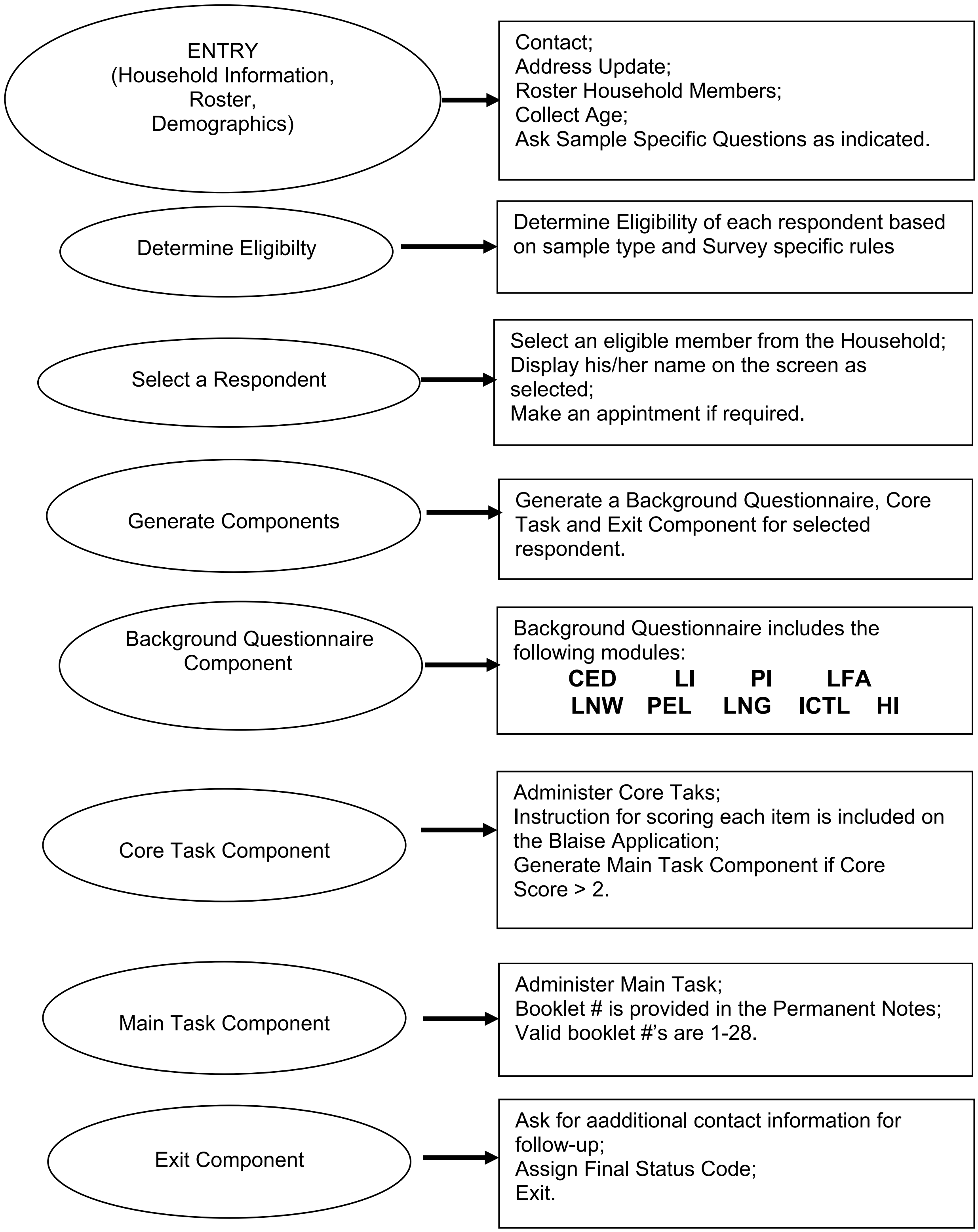
In IALSS, a respondent had to complete the background questionnaire, pass the core block of literacy tasks, and attempt at least five tasks per literacy scale in order for researchers to be able to estimate his or her literacy skills directly. Literacy proficiency
data were imputed for individuals who failed or refused to perform the core literacy tasks and for those who passed the core block but did not attempt at least five tasks per literacy scale. Because the model used to impute literacy estimates for non-respondents relies on a full set of responses to the background questions, IALS countries were instructed to obtain at least a background questionnaire from sampled individuals. They were also given a detailed non-response classification to use in the survey.

Each country was responsible for hiring its own interviewing staff. Thus, the number of interviewers, their pay rates, and the length of the survey period varied among the countries according to their norms and budgets. Each country was provided with a booklet to be used in training interviewers.

In Canada, the IALSS was collected by experienced Statistics Canada interviewers using Computer Assisted personal Interviewing technology.

The diagram on the next page graphically depicts the design of the International Adult Literacy and Skills Survey.

International Adult Literacy and Skills Survey
Main Collection Design



6.3 Scoring

Respondents' literacy proficiencies were estimated based on their performance on the cognitive tasks administered in the assessment. Unlike multiple-choice questions, which are commonly used in large-scale surveys and which offer a fixed number of answer choices, open-ended items such as those used in the IALSS elicit a large variety of responses. Because raw data is seldom useful by itself, responses must be grouped in some way in order to summarize the performance results. As they were scored, responses to the IALSS open-ended items were ultimately classified as correct, incorrect, or omitted.

The models employed to estimate ability and difficulty are predicated on the assumption that the scoring rubrics developed for the assessment were applied in a consistent fashion within and between countries. Several steps were taken to ensure that this assumption was met. Two of these main steps were the intra-country and inter-country rescores described in the following sections.

6.3.1 Intra-country rescoring

A variable sampling ratio procedure was set up to monitor scoring accuracy. At the beginning of scoring, almost all responses were rescored to identify inaccurate scorers and to detect unique or difficult responses that were not covered in the scoring manual. After a satisfactory level of accuracy was achieved, the rescoring ratio was dropped to a maintenance level to monitor the accuracy of all scorers. Average agreements were calculated across all items. To ensure that the first and second scores were truly independent, certain precautions had to be taken. For example, scorers had to be different persons, and the second scorer could not be able to see the scores given by the first scorer.

Scorers who received identical training within a country are expected to be more consistent with one other than with scorers in other countries. This expectation was confirmed during the 1994 IALS. Most of the rescoring reliabilities were above 97 percent. It is important to note that the results were well within the statistical tolerances set for the study and considerably better than those realized in other large-scale studies using open-ended items.

Table 6.1 displays the intra-country reliability from the 2003 IALSS in Canada.

Table 6.1 Intra-country scoring per cent reliability by domain

	Domaine			
	Prose and document	Numeracy	Problem solving	Total
		%		
Canada English scoring Canada French	95	95	92	95
Canada French scoring English	95	97	94	95

Source: International Adult Literacy and Skills Survey, 2003.

Since intra-country rescoring was used as a tool to improve data quality, score updates were not made to the database. In other words, the agreement data presented here indicate the minimum agreement achieved in scoring. After intra-country reliabilities were calculated, a few

scorers were found to be unreliable. These scorers either received additional training or were released. Where scores and rescores differed, the first scores were replaced with correct scores if the inaccuracy was due to a systematic error on the part of the first scorer. In some cases, the scoring guide was found to be ambiguous. In such cases, the scoring guide was revised and the first scores were changed to reflect the revisions, but the second scores were not altered. The second scores were never replaced, even if they were subsequently found to be erroneous.
In sum, the first scores reflect changes and corrections resulting from lessons learned in the intra-country rescoring analysis. The first scores are therefore more accurate and consistent than the second scores, which retain errors and thereby underestimate the rescore reliabilities somewhat. The extent to which the reliabilities are underestimated must be very small, however, given that most of the reliabilities are above 97 percent. These values indicate that very consistent scoring was achieved by all the participating countries.

6.3.2 Inter-country rescoring

Even after ensuring that all scorers were scoring consistently, fixing ambiguities in the scoring guides, and correcting any systematic scoring errors, it was still necessary to examine the comparability of scores across countries. Accurate and consistent scoring within a country does not necessarily imply that all countries are applying the scoring guides in the same manner. Scoring bias may be introduced if one country scores a certain response differently from the other countries. The inter-country rescorings described in this section were undertaken to ensure scoring comparability across countries.

As noted earlier, responses to the ALL assessment items were scored by each country separately. To determine inter-country scoring reliabilities for each item, the responses of a subset of examinees were scored by two separate groups. Usually, these scoring groups were from different countries. For example, a sample of test booklets was scored by two groups who scored Canada/English booklets and United States booklets. Inter-country score reliabilities were calculated by Statistics Canada, then evaluated by ETS. Based on the evaluation, every country was required to introduce a few minor changes in scoring procedures. In some cases, ambiguous instructions in the scoring manual were found to be causing erroneous interpretations and therefore lower reliabilities.

Using the inter-country score reliabilities, researchers can identify poorly constructed items, ambiguous scoring criteria, erroneous translations of items or scoring criteria, erroneous printing of items or scoring criteria, scorer inaccuracies, and, most important, situations in which one country consistently scored differently from another. In the latter circumstance, scorers in one country may consistently rate a certain response as being correct while those in another country score the same response as incorrect. This type of score asymmetry must be eliminated before the IRT scaling is performed. ETS and Statistics Canada identified such items, while the country in which the scoring problem occurred investigated the plausible causes for such systematic bias in scores. Where a systematic error was identified in a particular country, the original scores for that item were corrected for the entire sample.

Table 6.2 summarizes the inter-country rescore reliabilities for the Canadian component of the ALL study before corrections.

Table 6.2 Inter-country rescore reliability results

	Psychometric domain			
Country pairing (rescoring country – original country)	Prose and document (%)	Numeracy (%)	Problem solving (%)	Total (%)
Canada English – Canada French	95	95	92	95
Canada French – Canada English	95	97	94	95
Norway – Canada	91	93	91	92
Canada – United States	94	97	...	95
United States – Canada	95	97	...	95
United States – Bermuda	91	94	...	90
Bermuda – United States	93	95	...	93
Canada French – Switzerland	95	98	97	96
Switzerland – Canada French	94	96	94	95
Switzerland – Italy	96	98	96	96
Italy – Switzerland	93	97	93	94
Canada – Bermuda	...	...	83	83
Canada – Nuevo Leon, Mexico	91	95[1]	...	92

... Not applicable.
1. Quantitative literacy.

6.4 Data capture, data processing and coding

As a condition of their participation in the ALL, the IALSS team was required to capture and process their files using procedures that ensured logical consistency and acceptable levels of data capture error.

While the Canadian Background Questionnaire was collected on a computer, the task booklets needed to be scored and data captured, and, as part of the ALL guidelines, countries were advised to conduct complete verification of the captured scores (i.e., enter each record twice) in order to minimize error rates. Because the process of accurately capturing the test scores is essential to high data quality, 100 percent keystroke validation was performed.

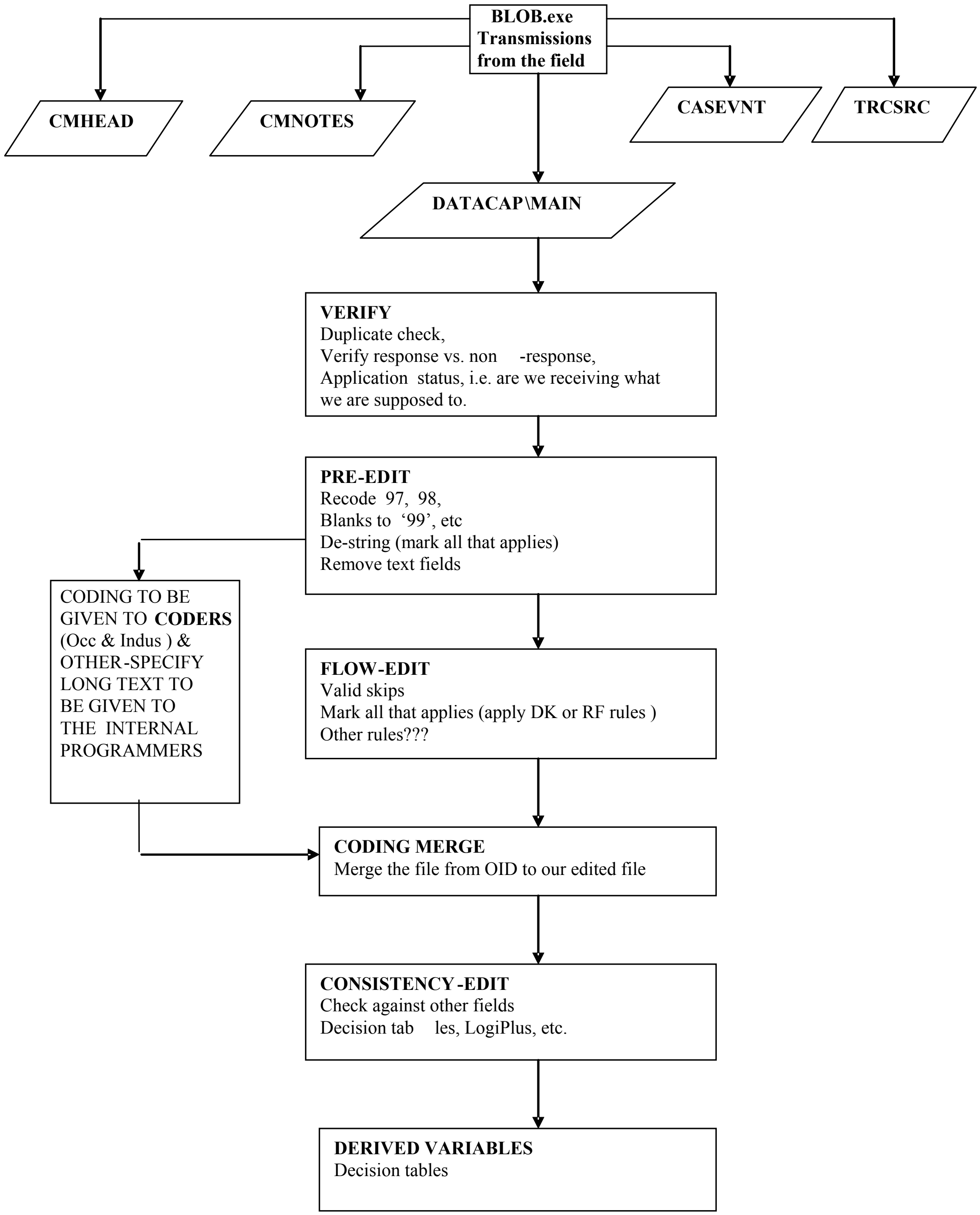
Each country was also responsible for coding industry, occupation, and education using standard international coding schemes (International Standard Industrial Classification, or ISIC; International Standard Occupational Classification, or ISOC; and International Standard Classification of Education, or ISCED). Further, coding schemes were provided for open-ended items, and countries were given specific instructions about the coding of such items so that coding error could be contained to acceptable levels.

In addition to the the Canadian file was also coded to National classifications for education, occupation (SOC, NOC and NOCS) and industry (SIC and NAICS). While these Canadian codes could not be included in the public microdata file due to specific industry and occupation code overlaps, special tabulations using these alternate classifications are available from Statistics Canada.

To create a workable comparative analysis, each IALSS country was required to map its national dataset into a highly structured, standardized record layout. In addition to specifying the position, format, and length of each field, this International Record Layout included a description of each variable and indicated the categories and codes to be provided for that variable.

The following technical flowchart cryptically outlines the steps taken to process the Canadian data file as it arrived from the field.

IALSS PROCESSING FLOW CHART



As illustrated above, the process can best be described as the culmination of 9 distinct steps:

1. Post-Collection Processing
2. Clean-up
3. Pre-Edit
4. Flow-Edit
5. Consistency Edit
6. Derived variable creation and imputation
7. Master file creation
8. National Public Use Microdata File creation
9. International Public Use Microdata File creation

6.4.1 Post-Collection Processing (Step 1)

The IALSS used three principal data collection methods: A Paper and Pencil Interview (PAPI), a Computer Assisted Interview (CAPI) and Computer Assisted Telephone Interview (CATI). The post-collection steps for CATI, CAPI and PAPI are provided in this section.

The Collection step, gathered Screener information about the Respondent and other members of the Household (i.e. the demographics: age, sex, immigration status, mother-tongue, etc, confirmation of address and eligibility of Household members), when telephone number of a given address were available. The respondent was randomly selected from the eligible members of the household. The interviewer gathered the required information through CATI, with the guidance of Hard and Soft appointments, at which time a personal visit was required in the language of their choice. Part of this interview (the Background Information) will be performed through CAPI application and part was performed through PAPI (the Core and the Main). If no contact had been made or there was no telephone number available for a given address, then a personal visit to the Household is required, applying the same procedures, but this time the Screener was done in person, rather than by telephone.

Running the BLOB

The BLOB process is simply a process to dismantle data into various file structures as established by the record-layouts defined by the programmers who developed the CAPI Computer Assisted Interviewing Application. The output of this process are a series of files such as the Header, Background questionnaire, casevnt, and other related files to perform required for processing.

The record-layouts provided by the CAPI programmers were used to create the Database files structures and layouts to be used for further processing. A processing unit ensured that the layouts created by the programmers were correct with meaningful and unique variable names that reference the data in the correctly assigned positions and have been assigned proper lengths and type of variable identified (i.e. numeric verses character).

6.4.2 The Clean-up Process (Step 2)

The purpose of the Clean up step is to identify any case that requires some type of intervention and identify cases that will not require editing because they are non-response. The process is intended to make corrections at the record level, such as identifying and resolving duplicate records, identifying and finding missing records, reviewing status code assignment and identify non-response cases.

Duplicate records

For the IALSS, only one record per sample unit can be included on a file. All duplicate records were thus either resolved or purged.

There were two types of duplicate records – full duplicates and ID duplicates.

- Full duplicates are defined as two identical records with exactly the same information in all fields. This occurs mostly in a CAPI environment when, for example, a record is transmitted and because of some type of system failure is re-transmitted, resulting in two records for a same case. In a PAPI environment it is possible that a record will be data captured twice, especially if the data capture system does not have a built in Sample Id validation process.
- ID duplicates are defined as two records with the same Sample ID that contain different data. In CAPI, this typically occurs as a result of a case being transmitted prior to being finalized, and again, after the case is finalized. In CAPI, an incorrect Sample Id may have been captured, thus creating a duplicate. These duplicates need to be examined carefully by a member of the processing team, to determine if the record needs to be dropped or if an update is required to the Sample ID.

Verify BQ with sample file

The sample id captured in Survey Data File was checked against the sample id in the sample file send out for collection. Records received on daily basis were required to be a subset of the sample file. If such was not the case, an investigation was initiated and the appropriate corrective actions were taken. Reports were monitored on daily basis throughout the collection period.

Verify Final Status Codes (BQStatus)

Status codes received from data collection were verified based on the information captured. The following table provides a description of the various status codes:

Response

Description	Outcome code
Fully completed	70
Partially completed	71

Non-response

Description	Outcome code
Refusal (Respondent refused to complete the Background Questionnaire or someone else refused for the respondent)	**80**
Language difficulties (The interview cannot take place because the respondent cannot communicate in the language required for the interview and the interviewer is not able to find an interviewer that can speak the language of the respondent)	**22**
Reading & Writing or Physical Disability (The Background questionnaire is not completed because of the respondent's mental or physical condition).	**92**
Death Sickness or Unusual Circumstances (Illness, fire flood, etc, no responses provided on the BQ)	**90**
Not eligible (No members in the household aged 16 and older or members do not fulfill sample selection criteria needed to target specific sub-populations)	**56 /76, 40**
Dwelling related non-response (Not at home, Dwelling under construction etc.)	**51, 52, 53, 55, 26**
Unable to contact the respondent (No contact, no one home/no answer, absent for the duration of survey, collection period ended)	**10, 11, 20, 98**
Weather condition (Interview prevented due to weather condition)	**18**
Duplication –(Dwelling was selected twice for the survey	**63**

Identify missing records

Missing records were identified early in the process in order to facilitate their resolution.

Procedures

There are many ways to safeguard against loss of documents during the Survey process. Some of those used in the IALSS are:

- ❖ Using a Master Assignment Control (MAC) form (listing each interviewer and their assignment)
- ❖ Interviewer Transmittal Form – Interviewers were directed to return survey documents periodically, with every "transmission" or "document package" they were required to attach a Transmittal form, which includes information similar to that found in the MAC.
- ❖ When documents arrived in Ottawa, the content of the package was verified against the Interviewer transmittal form. Any missing documents were immediately reported to the interviewer for resolution.
- ❖ Comparing the Sample File to the Survey Data File – Once all documents/records have been received, data captured and duplicate records have been resolved, the Master Sample file created by the Methodologist working on the survey, can be compared to the Survey Data File (file at data capture) to identify missing records or components.

Creation of response and non-response files

Once all records had been data captured and all steps identified in this section had been executed and all issues and problems resolved, the response and non-response files were prepared for processing.

Response Data File

Using the Survey Data File, split off all records that have some data and create a new file which we will now refer to as the “Response Data File”. This file will be used throughout the pre-editing and editing steps described in this document.

Non-Response Data File

Using the Survey Data File split off all records that have no data. Ensure that there is no valid data present in this file. If so, take appropriate actions to merge necessary records to the response file.

Steps such as comparing the Master Sample File to the Survey Data File and creation of the Response Data File and Non-Response Data File were only completed once data collection was completed and all records had been data captured. However, some comparison were made on daily basis, by tracking questionnaires, as soon as they were returned to the the processing team. This not only improved the chances of finding any missing documents, but also helped evaluate quality and accuracy of data, in terms of response verses non-response and sample-file comparisons

6.4.3 The Pre-Edit (Step 3)

The Pre-Edit Process is preformed after the Clean-up step. It consists of modifying the data at the individual variable level. Variables may be dropped, re-coded, re-sized or left as is. Mark-all that apply fields are "De-strung". Text fields are stripped off the main files and moved to a separate file.

Coding Other Specify

Questions that contain a list of answer categories often contain "*other"* as the final category. Coding these fields consisted of verifying the content and making a decision on whether the answer should have been included in another answer category or if they should stay in the “other” category.

This process was done during the pre-edit stage, this meant that the ‘write-in’ response were data captured and the values reviewed as part of a coding process.

If caught early in the data collection stage, systematic erros can be identified and the interviewers can be instructed to look closely at the answer categories and make themselves more aware of different answers that can be coded to given answer categories. This type of feedback during the process helps minimize the post collection processing required and, of course, increases the quality of the collected information.

In some cases the information in the "*other specify"* can be quite useful, such as in question D1 of the BQ. The following example illustrates a case where the information entered in the “Other

Specify" field would be coded to one of the existing categories. In this case to "<01>", for he/she is on work status, but is on Paternity Leave.

LFA_QD1

I would now like to talk about your employment status. What is your <u>current</u> work situation? Are you now...

<01> **employed or self employed?**
<02> **not working <u>and</u> looking for work?**
<03> **retired?**
<04> **student (including work programs)?**
<05> **doing unpaid household work?**
<06> Other – Specify……………………………………………………go to LFA_SD1
<98> Refused
<99> Don't know
Default Next Question: **LFA_QD2**

LFA_SD1

What is your current work situation?
Paternity Leave

For the IALSS survey no additional codes were to be added to the international variables. The following questions in the BQ contain "other Specify" fields. Their respective answers are in the appropriate questions identified by the arrow.

- QA0B – "Ethnic or cultural group do your ancestors belong?" → SA0B
- QA1D – "Country in which you were born…" → SA1D
- QA1E – "Current Immigrant Status…" → SA1E
- QA6D – "Country in which the highest level of education was attained, if that education was attained outside Canada…" → SA6D
- QB1 – "Language first learned in childhood…" → SB1
- QB2 – "Language most spoken at home…" → SB2
- QD1 – "Employment Status …" → SD1
- QF30H – Reasons for not taking education…→SF30F
- QH3H – Other purposes for using a computer.. → SH3H
- QH5N – Other purposes for using the Internet →SH5N

Destringing Mark all the Apply

Questions that allow multiple reponses for each respondent must be broken into single binary variables for each allowable choice

In the example below, QF14 is broken into 6 variables, one for each legitimate response category.

Example of De-string using QF14 - Who contributed towards the direct expenses…

Question	Original Answer Value	New Value Labels	New Variable name
QF14	<01> You personally		
		F14A	1 (Yes)
	<02> An employer		2 (No)
			7 (Don't Know)
	<03> A government agency		8 (Refused)
		F14B	1 (Yes)
	<04> A union or professional assoc.		2 (No)
			7 (Don't Know)
	<05> Anyone else		8 (Refused)
		F14C	1 (Yes)
	<06> No DIRECT EXPENSE		2 (No)
			7 (Don't Know)
	<98> Refused		8 (Refused)
		F14D	1 (Yes)
	<99> Don't know		2 (No)
			7 (Don't Know)
			8 (Refused)
		F14E	1 (Yes)
			2 (No)
			7 (Don't Know)
			8 (Refused)
		F14F	1 (Yes)
			2 (No)
			7 (Don't Know)
			8 (Refused)

If F14 were 99(Don't know) 98 (Refused) or blank, the program would have to code all of the new variables (F14A to F14F) to 97, 98 or 99 respectively.

Other pre-edit actions

Each variable passing through the pre-edit stage had one or more of the following actions taken as part of this processing step.

TERM	DEFINITION
Drop	variable is removed from the file
Re-code	Don't know and refusal are set to the proper codes (7 and 8), any other code that may have been different in collection is coded to comply with International Standards
Leave	No modification to the variable is required
Re-size	the length may need to be increased or decreased
De-String	Mark-all that apply questions are returned in a compressed fashion, and need to be converted into yes/no values, 1 per possible answer
Text	Written to Long answer file and removed from data file

The following is an example of a pre-edit specification using <u>various</u> sections of the Background questionnaire:

InName	Insize	InType	OutName	OutSize	Recode Flag	Value Labels	Comment
ID (sampleid)	9	C	ID	9	2		Leave
…							
BT1	5	N		0	0		Drop/Delete
B1 (01-14, 98)	4	N	B1A	2	5		Leave
			B1B	2	5		
B2 (01-14, 98)	2	N	B2	2	2		Leave
BT2	5	N		0	0		Drop/Delete
..							
D25	80	C	D25	1	3	1=yes 2=no 7=Don't know 8=Refused	Text and recode
D26	80	C	D26	1	3	1=yes 2=no 7=Don't know 8=Refused	Text and recode
D27	80	C	D27	1	3	1=yes 2=no 7=Don't know 8=Refused	Text and recode
D28	80	C	D28	1	3	1=yes 2=no 7=Don't know 8=Refused	Text and recode
..							
..							
ET1	5	N		0	0		Drop/Delete
E1A	1	N		1	2		Leave
E1B	1	N		1	2		Leave
E1C	1	N		1	2		Leave

6.4.4 Flow Edits (Step 4)

The Flow Edit replicates the flow pattern, (question order) intended by the survey design. The questionnaire includes a series of conditional branching depending on previously supplied answers. For instance, if the respondent indicates not ever taking any course or programs of study, the the entire adult education section is skipped. Fields that are skipped because of flows are converted to blanks (valid skip), as well fields that are skipped because of non-response are set to "9"'s (Not Stated). For the IALSS processing, a Top-down approach was used by all participating countries. The top-down approach follows the flow of the questionnaire exactly starting with the first question, therefore if the answer to the first question requires a skip of 3 questions then the 3 questions skipped would be set to blank without regard to their contents.

The following is a flow edit example for questions QA4 D to QA8, the same type of specifications should be developed for each section within the BQ.

Check Item	Variable	Size	Type	Condition	Goto	Value	Comments
	QA4D – "Have you ever taken but not completed any schooling towards a trade/vocational, college, CEGEP or university certificate, diploma or degree?"	1	N	1			
	QA4D	1	N	2	QA6	VS	Fill AQA4E with **6**, QA4F with **96**, QA5 with **96**
	QA4D	1	N	8 (RF), 9(DK)	QA6	NS	Fill AQA4E with **9**, QA4F with **99**, QA5 with **99**
	QAE – "Was the CEGEP diploma or certificate part of a university transfer program?	1	N				No action needed, since the whole question goes to the next question (QA4F)
	QA4F – "What was the normal length of time required to complete this certificate or diploma when taken full-time?	2	N	01, 02, 03, 04, 05, 06, 07	QA6	VS	Fill QA5 with **96**
	QA4F	2	N	98 (RF), 98(DK)	QA6	NS	Fill QA5 with **99**
	QA5 – "How old were you when you took this schooling towards a trade/vocational, college, CEGEP or university certificate, diploma or degree?"	2	N	0 - 95	CA6A	VS	Fill QA6 with **96,**
	QA5	2	N	98 (RF), 98(DK)	CA6A	NS	Fill QA6 with **99**
	QA6 - "How old were you when you completed your highest level of schooling?"	2	N				No action needed, since the whole question goes to the next question (QA4F)
C	CA6A			If (QA1 = 2 or IMM_Q01 = 2) AND QA3B > 0 AND QA3B	QA6B	VS	
				ELSE	QA6E	VS	

Check Item	Variable	Size	Type	Condition	Goto	Value	Comments
	OR						
C	CA6A			If NOT((QA1 = 2 or IMM_Q01 = 2) AND QA3B > 0 AND QA3B <96)	QA6E	VS	The else of this is understood to go to the next question QA6B
	QA6B – " Did you obtain this level of education in Canada?"						No action needed, since the whole question goes to the next question (QA4C)
	QA6C –"What was the highest level of education you obtained outside of Canada?"						No action needed, since the whole question goes to the next question (QA4D)
	QA6D – "In what country did you obtain this level of education?"						No action needed, since the whole question goes to the next question (QA4E)
	QA6E –"What was the language used most often for teaching your courses at elementary school, excluding language courses?"						No action needed, since the whole question goes to the next question (CA7)
C	CA7			If QA4B = 1 or QA4B =2	QA10	VS	The else of this is understood to go to the next question QA7A
	QA7A – "What was the language used most often for teaching your courses at high school or junior high school, excluding language courses?"						No action needed, since the whole question goes to the next question (QA8)
	QA8 – "In what province or territory were you last in high school or junior high school?"			10, 11, 12, 13, 35, 46, 47, 48, 59, 60, 61, 62, 76, 77	QA9A	VS	Fill QA8B with **6**
	QA8			24	CA8A		
C	CA8A			QA4A NE 1	QA9A	VS	Fill QA8B with **6**
	QA8B						No action needed, since the whole question goes to the next question (QA9A)
	QA9A						

6.4.5 Consistency Edits (Step 5)

Consistency Edits examines the coherence of information between different variables. Many of the consistency edits were built into the IALSS CAPI application, so values that do not concur with previous information was often flagged by the interviewer and clarified at the source. For the Canadian files, the actual steps required to conform to the international standards were minimal. However, all of the required edits were repeated during Canadian processing to ensure data quality. This step was done immediately following the flow edits process.

Each countries defined their own tolerance levels for many of these questions depending on what was reasonable for their country (e.g. upper and lower boundaries for number of years required to complete a program, upper and lower boundaries for income levels, etc).

For the Canadian IALSS, a breach of a tolerance level resulted in a prompt by the interviewer for clarification. However, in some instances, such breaches could only be corrected after the collection. This would, of course, require corrective action that would be independent of the respondent. Hence the need for a minimum number of consistency edits.

The following consistency edits were performed by the IALSS processing team to satisfy the ALL processing needs.

Consistency Edits related to Age of Respondent

Edit#	Question	Rule
E1	Age (derived from DOB)	Age at date of interview is 16 or over **((DOBMM > CURRMTH) AND DOBYY=1987)** OR (**E0 <16 or Agerange=1) then E1**
E2	QA2 – Year of immigration	Year of immigration cannot be less than year of birth **QA2 < DOBYY AND (QA1E =1 OR QA1E=2) AND (QA1=2 OR IMM_Q01=2) then E2**
E3	QA3 - # of years of formal education	Verify # years of education (A3) and level attained (A4) are valid. **Examine the relationship between QA3 and (QA4A-QA6C) and ensure that there is no conflict between years of education and education attained. If conflict then E3**
E4	QA5 – Age when took schooling towards…	Verify that Age entered is not Greater than Age of respondent **QA5 >E0 AND QA5 NE 96 then E4**
E5	QA6 – Age when complete <highest level of education>	Verify that Age entered is not Greater than Age of respondent **QA6 >E0 AND QA6 NE 96 then E5**

Consistency Edits related to Employment

<table>
<tr><th>Edit#</th><th>Question</th><th>Rule</th></tr>
<tr><td>E6</td><td>QD3A- Last worked</td><td>Year entered cannot be in the last twelve months
QD3B||QD3A > CurrYear-1||CurrMth then E6</td></tr>
<tr><td>E7</td><td>QD5MA-QD5ML – Months worked
QD6AMA-QD6AML
– Worked whole mth
QD6BMA-QD6BML
– Worked Part of the month</td><td>Months entered in QD6AMA-QD6AML & QD6BMA-QD6BML cannot be any other months than those entered in QD5MA-QD5ML respectively
(QD5MA-QD5ML) = 1 and (QD6AMA-QD6AML) ne 1 and (QD6BMA-QD6BML) ne 1 respectively then E7</td></tr>
<tr><td>E8</td><td>QD6AMA-QD6AML
– Worked whole month
QD6BMA-QD6BML
– Worked Part of the month</td><td>Months reported in these two questions must be mutually exclusive.
(QD6AMA-QD6AML) = (QD6BMA-QD6BML) respectively then E8</td></tr>
<tr><td>E9</td><td>QD9 – Ever without work …</td><td>Must have answered yes in QD4 (worked every month in reference period)
IF response in QD9 and QD4 NE 1 then E9
i.e. QD4 NE 1 AND QD9 IN (1,2,8,9)</td></tr>
<tr><td>E10</td><td>QD21A –QD21B
– When start working at this job…</td><td>Date entered cannot be later than interview date
QD21B || QD21A > CurrYear || CurrMth then E10</td></tr>
<tr><td>E11</td><td>QD24A-QD24B – When last worked at this job…</td><td>Date entered must be => QD21A-QD21B
QD24B || QD24A < QD21B || QD21A then E11</td></tr>
<tr><td>E12</td><td>QD40 - # Days per week</td><td>Must have answered Per Day (2) in question QD39
If response in QD40 = response and QD39 NE 02 then E12</td></tr>
</table>

Consistency Edits related to Wages and Salary

Edit#	Question	Rule
E13	QD41 – Wages	Amount entered must be appropriate to the answer in QD39, if not then E13 **Example QD39 =01 (Per hour) - acceptable range - $1 to $1000** **QD39 =02 (Per day) - acceptable range - $50 to $600** **QD39 =03 (Weekly) – acceptable range - $25 to $2000** **QD39 =04 (Bi-weekly) – acceptable range - $50 to $3000** **QD39 =05 (Twice a month) – acceptable range - $50 to $3000** **QD39 =06 (Per month) – acceptable range - $150 to $8000** **QD39 =07 (Per year) – acceptable range - $1500 to $100000** **QD39 =08 (Per piece) – acceptable range - $1 to $5000**
E15	QD42 - Wages/month	Amount entered must be appropriate for monthly wages, if not then E15 **QD39 = Per month – acceptable range - $150 to $8000**

Consistency Edits related to Program of study

Edit#	Question	Rule	Comments
E16	QF8 – Weeks spent in program	Cannot be greater than 52 (Tolerance level only) **QF8 >52 and QF8 < 96 then edit E16**	OK
E17	QF9 – Avg # hours/week	Should not be greater than 30 (Tolerance level only) **QF9 >30 and QF9 < 96 then edit E17**	Tolerance level should be changed to >40, for there are 98 cases, that have a value of 40 and 31 cases between the values of 30 and 40.
E18	QF11 – hours/day	Should not be greater than 6 (Tolerance level only) **and QF11 > 6 and QF11 < 96 then edit E19**	Tolerance level should be changed to >8, for there are 184 cases that have a value of 8 .

Consistency Edits related to Activities outside of school

Edit#	Question	Rule
E19	QG5A to G5H – Parts of newspaper K11	If response in QG3A is equal 1,2 or 3 then G5A to G5H should be answered **If QG3A in (1,2,3) then do;** **if not(QG5A =1 or QG5B=1 or QG5C =1 or QG5D=1 or QG5E =1 or QG5F=1 or QG5G =1 or QG5H =1) then E25;** **end;**

Consistency Edits related to Total Income

Edit#	Question	Rule
E20	K11	Income reported must be => K6 (Personal Income) **If (. <Qk6 <9999996.96) and (. < Qk11 < 999999.96) then do;** **If QK11 < QK6 then E28;** **end;**

It should be pointed out that no other consistency edits were performed on the Canadian file, and so, there likely remains cases were inconsistencies persist (such as between the wages information reported in Section D and the Income information reported in Section K). The most appropriate treatment for these outliers is left to the individual researcher.

6.4.6 Imputations and derived variables (Step 6)

International rules dictated that imputations were required on some key variables, namely, Age, Agegroup, education Level, year of immigration, aboriginal and mother tongue. However, in Canada the imputations were done on only five variables; mother tongue was excluded from imputation because the field was accurately recorded for every Canadian respondent. A number of derived variables were created to aid research into the antecedents and outcomes of skills. In most cases, these have not been retained in the Public Use Microdata file. However, those that have been appended to the Master file are described in details in the section following the description of imputations steps.

Imputation

Generally speaking, the imputation step offers plausible data in the place of missing values, however, in some cases, the screener information and some other related information, helped clarify and alter some of the inconsistencies in the values of these key variables. Such imputations were typically performed immediately following the consistency edits though since some of the inconsistencies were dealt with in the preceding edit steps and thus, imputation was kept to a minimum.

Age of respondent

a) Primary source: age calculated from date of birth of respondent in BQ.

b) If calculated age is missing, then age reported in the screener, was used, if it was consistent with the age group reported in the BQ. (There were 131 such cases).

c) If the age was not available from sources a) and b), then donor imputation was used: randomly selected a donor with matching age group. (There were 21 such cases)

Education Level

Only 4 response records had a missing education level. The imputation model used values to a host of available questions such as years of schooling, adult education experience and, in one case, a hot deck method using a suitably matched donor was used to provide this crucial missing piece of information.

Born in Canada

Question QA1 (born in Canada) was not asked in the BQ if the same question was already asked in the screener, to the respondent. Missing values like Refused and Don't know in QA1 were imputed by the value in the screener. All cases, therefore, had a response in QA1 after this imputation.

Year of Immigration

Question QA2 (year immigrated to Canada) is required to distinguish between recent and established immigrants. Missing values in this variable was imputed.

a) Year of immigration set as follows: Current year minus years lived in Canada. YearImmig = 2003 – QA2B (years lived in Canada). (There were 7 such cases).

b) A randomly selected donor with matching country of birth and age group. Consistency was ensured with the edit rule: year of birth <= year of immigration. (There were 2 such cases)

Aboriginal Status

The aboriginal question was asked of all respondents. However, the BQ Aboriginal status question was not asked if the same question was already asked in the screener to the respondent. The aboriginal question is a “mark all that apply” question. Only Manitoba, Saskatchewan, and the three territories were imputed.

a) Missing values like Refused and Don’t knows in the Aboriginal question was imputed by response in the screener. (There were 3 such cases).

b) There were 9 cases with inconsistent responses (e.g. <1> Yes, North American Indian, and <4> Not Aboriginal). These cases were resolved manually by examining responses to ethnic origin, immigration, and mother tongue. (There were 10 such cases)

Mother Tongue

Question QB1 - Mother tongue was another variable that had to be imputed in order to conform to the international standards. In Canada, since there were no missing values in this field; no imputation was required for mother tongue.

Derived variables

Another important step regards deriving summary variables based on the detailed information provided by the respondents. The following list of derived variables are included on the IALSS Master Data file. Please note, these derived variables are not included on the National or International IALSS Public Use Microdata Files for Canada. The descriptions, is provided for information purposes and include the logic used to create the derived variable on the Master Data file.

Simple derived variables

Labour force activities

Labour force participant at time of survey (2 categories)

DV name:	LFP1
DV label:	Labour force participant at time of survey (2 categories)
DV value labels:	0 'Not in labour force'; 1 'In labour force'
Source question(s):	D1: What is your current work situation?
DV pseudo logic:	D1 identifies whether respondent was in the labour force at time of survey; that is either employed, self employed or unemployed and looking for work.

Labour force participant at time of survey (3 categories)

DV name:	LFP2
DV label:	Labour force participant at time of survey (3 categories)
DV value labels:	0 'Not in labour force'; 1 'Unemployed'; 2 'Employed'
Source question(s):	D1: What is your current work situation?
DV pseudo logic:	D1 identifies whether respondent was in the labour force at time of survey; and among those in labour force distinguishes between those employed and unemployed.

Work pattern in last 12 months

DV name:	LFSTAT12
DV label:	Work patter in last 12 months
DV value labels:	1 'Employed all year'; 2 'Unemployed all year'; 3 'Not in labour force all year'; 4 'Employed part-year, unemployed part-year'; 5 'Employed part-year, not in the labour force part-year', 6 'Unemployed part-year, not in the labour force part-year', 7 'Employed, unemployed, and not in labour force part-year'
Source question(s):	D2: Did you work at a job or business at any time in the last 12 months? D4: In last 12 months, did you work in every month (at least eight hours in each month)? D5: In which months did you work (at least eight hours in each month)? D6A: In which months did you work the whole month, that is, at least eight hours every week? D6B: In which months did you work part of the month? D9: Were you ever without work for a week or more at any time during the last 12 months because you changed employers? D10: In which month(s) were you without work for a week or more because you were changing employers? D11: Were you ever without work for a week or more at any time during the last 12 months because you were on temporary layoff? D12: In which month(s) were you on temporary layoff? D13: Did you look for work at anytime during the last 12 months; that is, looking for a new or different job? D14: In which month(s) did you look for work?

DV pseudo logic: Question D2, D4, D5 *(mo1 to mo12)*, D6A *(mo1 to mo12)*, D6B *(mo1 to mo12)*, D9, D10 *(mo1 to mo12)*, D11, D12 *(mo1 to mo12)*, D13, and D14 *(mo1 to mo12)* are used to create 12 interim variables EMPMO1 to EMPMO12 to classify the employment status of respondents in each of last 12 months. The categories of these interim variables are 1 'Employed all month'; 2 'Unemployed all month'; 3 'Not in the labour force all month'; 4 'Employed part-month, unemployed part-month'; 5 'Employed part-month, not in the labour force part-month'. Information from these 12 interim variables is used to derive a 12 month work pattern variable, LFSTAT12.

Number of unemployment spells in last 12 months

DV name: UNEMP1
DV label: Number of unemployment spells in last 12 months
DV value labels: 1 '0 spells'; 2 '1 spell'; 3 '2 spells'; 4 '3 spells'; 5 '4 spells'; 6 '5 spells'
Source question(s): LFSTAT12 (dv): Work patter in last 12 months
EMPMO1 (dv): Work status in month 1
EMPMO2 (dv): Work status in month 2
EMPMO3 (dv): Work status in month 3
EMPMO4 (dv): Work status in month 4
EMPMO5 (dv): Work status in month 5
EMPMO6 (dv): Work status in month 6
EMPMO7 (dv): Work status in month 7
EMPMO8 (dv): Work status in month 8
EMPMO9 (dv): Work status in month 9
EMPMO10 (dv): Work status in month 10
EMPMO11 (dv): Work status in month 11
EMPMO12 (dv): Work status in month 12
DV pseudo logic: Derived variable LFSTAT12 and EMPMO1 to EMPMO12 are used to calculate the number of spells of unemployment experienced by respondents in last 12 months. An assumption is made in months with two labour force statuses. In this case the proportion allocated to each status is one-half-month. The year is defined as consisting of 24 part-months. Further, an assumption is made about the sequence of activities in months with two labour force statuses. The sequencing is done so as to minimize the number of times an individual changes from one state to another and to maximize the length of time an individual remained in any given status. This is first applied to employment statuses, and then to unemployment statuses.

Number of consecutive weeks unemployed at time of survey (weeks)

DV name:	UNEMP2
DV label:	Number of consecutive weeks unemployed at time of survey (weeks)
DV value labels:	Number of weeks
Source question(s):	LFP2 (dv): Labour force participant at time of survey (3 categories) LFSTAT12 (dv): Work patter in last 12 months EMPMO1 (dv): Work status in month 1 EMPMO2 (dv): Work status in month 2 EMPMO3 (dv): Work status in month 3 EMPMO4 (dv): Work status in month 4 EMPMO5 (dv): Work status in month 5 EMPMO6 (dv): Work status in month 6 EMPMO7 (dv): Work status in month 7 EMPMO8 (dv): Work status in month 8 EMPMO9 (dv): Work status in month 9 EMPMO10 (dv): Work status in month 10 EMPMO11 (dv): Work status in month 11 EMPMO12 (dv): Work status in month 12
DV pseudo logic:	Derived variable LFP2, LFSTAT12 and EMPMO1 to EMPMO12 are used to calculate the number of consecutive weeks unemployed at time of survey.

Number of consecutive weeks unemployed at time of survey (2 categories)

DV name:	UNEMP3
DV label:	Number of consecutive weeks unemployed at time of survey (2 categories)
DV value labels:	0 'Short term unemployed - less than 6 months'; 1'Long term unemployed - 6 months or more'
Source question(s):	UNEMP2 (dv): Number of consecutive weeks unemployed at time of survey (weeks)
DV pseudo logic:	UEMP2 is recoded into two categories.

Number of weeks unemployed in last 12 months

DV name:	UNEMP4
DV label:	Number of weeks unemployed in last 12 months
DV value labels:	Number of weeks
Source question(s):	EMPMO1 (dv): Work status in month 1 EMPMO2 (dv): Work status in month 2 EMPMO3 (dv): Work status in month 3 EMPMO4 (dv): Work status in month 4 EMPMO5 (dv): Work status in month 5 EMPMO6 (dv): Work status in month 6 EMPMO7 (dv): Work status in month 7 EMPMO8 (dv): Work status in month 8 EMPMO9 (dv): Work status in month 9 EMPMO10 (dv): Work status in month 10 EMPMO11 (dv): Work status in month 11 EMPMO12 (dv): Work status in month 12

DV pseudo logic: Derived variables EMPMO1 to EMPMO12 are used to calculate the number of weeks of unemployment in last 12 months. Monthly labour force statuses are converted into weeks as followed: first, the year is defined as consisting of 24 part-months; and second a part-month is converted to weeks based on the following calculation: 365.25/ (7*24).

Number of months without work for a week or more because of layoff

DV name: UNEMP5
DV label: Number of months without work for a week or more because of layoff
DV value labels: Number of months
Source question(s): D4: During the last 12 months, did you work in every month (at least 8 hours in each month)?
D11: Were you ever without work for a week or more at any time during the last 12 months because you were on temporary layoff?
D12: In which month(s) were you on temporary layoff?
DV pseudo logic: D4, D11, and D12 *(mo1 to mo12)* determine the number of months without work for a week or more for those who were employed in all 12 months. Respondents who worked in last 5 years must answer D29, and D29 is recoded and used to exclude those who did not work in last 5 years.

Number of months looking for work at anytime during the last 12 months

DV name: LK4WORK
DV label: Number of months looking for work at anytime during the last 12 months
DV value labels: Number of months
Source question(s): D13: Did you look for work at anytime during the last 12 months; that is, looking for a new or different job?
D14: In which month(s) did you look for work?
DV pseudo logic: D14 determines the number of months looking for work at anytime during the last 12 months.

Number of weeks not in labour force in last 12 months

DV name: NLFWKS1
DV label: Number of weeks not in labour force in last 12 months
DV value labels: Number of weeks
Source question(s): EMPMO1 (dv): Work status in month 1
EMPMO2 (dv): Work status in month 2
EMPMO3 (dv): Work status in month 3
EMPMO4 (dv): Work status in month 4
EMPMO5 (dv): Work status in month 5
EMPMO6 (dv): Work status in month 6
EMPMO7 (dv): Work status in month 7
EMPMO8 (dv): Work status in month 8
EMPMO9 (dv): Work status in month 9
EMPMO10 (dv): Work status in month 10
EMPMO11 (dv): Work status in month 11
EMPMO12 (dv): Work status in month 12
DV pseudo logic: Derived variables EMPMO1 to EMPMO12 are used to calculate the number of weeks not in the labour force in last 12 months. Monthly labour

force statuses are converted into weeks as followed: first, the year is defined as consisting of 24 part-months, and second a part-month is converted to weeks based on the following calculation: 365.25/ (7*24).

Number of consecutive weeks not in the labour force at time of survey

DV name: NLFWKS2
DV label: Number of consecutive weeks not in labour force in last 12 months
DV value labels: Number of weeks
Source question(s): LFP2 (dv): Labour force participant at time of survey (3 categories)
LFSTAT12 (dv): Work patter in last 12 months
EMPMO1 (dv): Work status in month 1
EMPMO2 (dv): Work status in month 2
EMPMO3 (dv): Work status in month 3
EMPMO4 (dv): Work status in month 4
EMPMO5 (dv): Work status in month 5
EMPMO6 (dv): Work status in month 6
EMPMO7 (dv): Work status in month 7
EMPMO8 (dv): Work status in month 8
EMPMO9 (dv): Work status in month 9
EMPMO10 (dv): Work status in month 10
EMPMO11 (dv): Work status in month 11
EMPMO12 (dv): Work status in month 12
DV pseudo logic: Derived variable LFP2, LFSTAT12 and EMPMO1 to EMPMO12 are used to calculate the number of consecutive weeks unemployed at time of survey.

Number of weeks employed in last 12 months

DV name: EMPWKS1
DV label: Number of weeks employed in last 12 months
DV value labels: Number of weeks
Source question(s): EMPMO1 (dv): Work status in month 1
EMPMO2 (dv): Work status in month 2
EMPMO3 (dv): Work status in month 3
EMPMO4 (dv): Work status in month 4
EMPMO5 (dv): Work status in month 5
EMPMO6 (dv): Work status in month 6
EMPMO7 (dv): Work status in month 7
EMPMO8 (dv): Work status in month 8
EMPMO9 (dv): Work status in month 9
EMPMO10 (dv): Work status in month 10
EMPMO11 (dv): Work status in month 11
EMPMO12 (dv): Work status in month 12
DV pseudo logic: Derived variables EMPMO1 to EMPMO12 are used to calculate the number of weeks employed in last 12 months. Monthly labour force statuses are converted into weeks as followed: first, the year is defined as consisting of 24 part-months, and second a part-month is converted to weeks based on the following calculation: 365.25/ (7*24).

Number of consecutive weeks employed at time of survey

DV name: EMPWKS2
DV label: Number of consecutive weeks employed in last 12 months
DV value labels: Number of weeks
Source question(s): LFP2 (dv): Labour force participant at time of survey (3 categories)
LFSTAT12 (dv): Work patter in last 12 months
EMPMO1 (dv): Work status in month 1
EMPMO2 (dv): Work status in month 2
EMPMO3 (dv): Work status in month 3
EMPMO4 (dv): Work status in month 4
EMPMO5 (dv): Work status in month 5
EMPMO6 (dv): Work status in month 6
EMPMO7 (dv): Work status in month 7
EMPMO8 (dv): Work status in month 8
EMPMO9 (dv): Work status in month 9
EMPMO10 (dv): Work status in month 10
EMPMO11 (dv): Work status in month 11
EMPMO12 (dv): Work status in month 12
DV pseudo logic: Derived variable LFP2, LFSTAT12 and EMPMO1 to EMPMO12 are used to calculate the number of consecutive weeks employed at time of survey.

Number of weeks employed at current or last main job

DV name: EMPWKS3
DV label: Number of weeks employed at current or last main job
DV value labels: Number of weeks
Source question(s): D34: How many months in a year <did/do> you usually work at this job or business?
D35: <Did/do> you usually work every week of the month at this job or business?
D36: On average, how many weeks in a month <did/do> you usually work at this job or business?
DV pseudo logic: D34, D35, and D36 determine the number of weeks employed at current or last main job. If respondent answered work every week of the month in question D35, 4.33 is the estimated number of weeks employed in a month.

Number of hours employed at current or last main job

DV name: EMPHRS
DV label: Number of hours employed at current or last main job
DV value labels: Number of hours
Source question(s): D34: How many months in a year <did/do> you usually work at this job or business?
D35: <Did/do> you usually work every week of the month at this job or business?
D36: On average, how many weeks in a month <did/do> you usually work at this job or business?
D37: On average, how many hours per week <did/do> you usually work at this job or business?

DV pseudo logic: D34, D35, D36 and D37 determine the number of hours employed at current or last main job. If respondent answered work every week of the month in question D35, 4.33 is the estimated number of weeks employed in a month.

Employment intensity in last 52 weeks (6 categories)

DV name: EMPFULL1
DV label: Employment intensity in last 52 weeks (6 categories)
DV value labels: 1 'Employed full time last 52 wks'; 2 'Employed full time but < 52 wks b/c not in labour force part of year'; 3 'Employed full time but < 52 wks b/c unemployed part of year'; 4 'Employed part time/equally part & full time last 52 wks'; 5 'Employed part time/equally part & full time but < 52 wks b/c not in labour force part of year'; 6 'Employed part time/equally part & full time but < 52 wks b/c unemployed part of year'
Source question(s): D7: During the last 12 months, did you work mostly full-time, that is <fill with country standard> hours per week or more, or part-time, that is <less than fill with country standard> hours per week?
LFP2 (dv): Labour force participant at time of survey (3 categories)
LFSTAT12 (dv): Work patter in last 12 months
EMPWKS1: Number of weeks employed in last 12 months
DV pseudo logic: D7, LFSTAT12, LFP2, EMPWKS1 determine the employment intensity in last 52 weeks prior to the survey.

Self employed status

DV name: EMPSELF
DV label: Self employed status
DV value labels: 0 'Non self employed status at main job'; 1 'Self employed status at main job'
Source question(s): D29: What <was/is> your status at this job or business?
DV pseudo logic: D29 determines whether respondent is self employed at current/last main job or not.

Supervisory status

DV name: SUPERVIS
DV label: Supervisory status
DV value labels: 0 'Non-supervisory status at main job'; 1'Supervisory status at main job'
Source question(s): D29: What <was/is> your status at this job or business?
DV pseudo logic: D29 determines whether respondent has supervisory responsibilities at current/last main job or not. An assumption is made that self employed with employees would imply such responsibilities.

Firm size

DV name: FIRMSIZE
DV label: Firmsize
DV value labels: 1 'Less than 20'; 2 '20 to 99'; 3 '100 to 499'; 4 '500 to 999'; 5 '1000 and over'
Source question(s): D30: About how many persons <were/are> employed at the location where you work(ed)? Would it be less than 20, 20 to 99, 100 to 499, 500 to 999, or 1000 and over?

D31: <Did/does> <your employer> operate at more than one location in <country>?
D33: In total, about how many persons <were/are> employed at all locations? Would it be less than 20, 20 to 99, 100 to 499, 500 to 999, or 1000 and over?

DV pseudo logic: D30, D31 and D33 determine the size of the company respondent work(ed) at, including all locations.

Tenure (months)

DV name: TENURE1
DV label: Number of months working at the same job
DV value labels: Number of months
Source question(s): D1: What is your current work situation?
D2: Did you work at a job or business at any time in the last 12 months?
D3: When did you last work at a job or business?
D21: When did you start working at this job or business?
D23: Are you still working at this job or business?
D24: When did you last work at this job or business?
DV pseudo logic: D1, D2, D3, D21, D23 and D24 determine the number of months respondents have worked at the same job. This refers to current or main job.

Tenure status (4 categories)

DV name: TENURE2
DV label: Tenure (4 categories)
DV value labels: 1 'Established tenure - held in all of last 12 months'; 2 'Recent tenure - held less than last 12 months'; 3 'Tenure of previous job - employed at different job'; 4 'Tenure of previous job - not employed last 12 months'
Source question(s): TENURE1: Number of months working at the same job
D23: Are you still working at this job or business?
DV pseudo logic: D23 and TENURE1 determine when tenure was earned and whether it was throughout the last 12 months.

Participation in education and learning

Destination of E/T program taken in last 12 months (students)

DV name: PROGDES1
DV label: Destination of E/T program taken in last 12 months (students)
DV value labels: 1 'Student in E/T program, ISCED levels 0 - 3'; 2 'Student in E/T program, ISCED levels 4 & 5B'; 3 'Student in E/T program, ISCED levels 5A & 6'; 4 'Student in E/T program, in formal institution, DK level'; 5 'Student in E/T program, but not in formal institution'; 6 'Student in E/T program, DK if in formal institution'; 7 'Student participated in E/T, but not in program'; 8 'Student participated in E/T, but DK if in program'; 9 'Student, but did not participate in any E/T'; 10 'Not a student at time of survey'
Source question(s): D1: What is your current work situation? (Student is one of the choices.)
F1: Did you take any education or training in last 12 months?
F2: Did you take any courses as part of a program towards a certificate, diploma, or degree?

	F3: Was this program given by a school, college, or university? F4: What type of certificate, diploma, or degree were you taking this education or training towards? (Following ISCED97 levels.)
DV pseudo logic:	D1, F1, F2, F3, and F4 determine the destination of education and/or training program taken in last 12 months. This variable is only for program participants who were students at time of survey.

Destination of E/T program taken in last 12 months (all respondents)

DV name:	PROGDES2
DV label:	Destination of E/T program taken in last 12 months (all respondents)
DV value labels:	1 'In E/T program, ISCED levels 0 - 3'; 2 'In E/T program, ISCED levels 4 & 5B'; 3 'In E/T program, ISCED levels 5A & 6'; 4 'In E/T program, in formal institution, DK level'; 5 'In E/T program, but not in formal institution'; 6 'In E/T program, DK if in formal institution'; 7 'Participated in E/T, but not in program'; 8 'Participated in E/T, but DK if in program'; 9 'Did not participate in any E/T'
Source question(s):	F1: Did you take any education or training in last 12 months? F2: Did you take any courses as part of a program towards a certificate, diploma, or degree? F3: Was this program given by a school, college, or university? F4: What type of certificate, diploma, or degree were you taking this education or training towards? (Following ISCED97 levels.)
DV pseudo logic:	F1, F2, F3, and F4 determine the destination of education and/or training program taken in last 12 months. This variable is for all respondents who reported participating in programs.

Participated in E/T program taken in last 12 months

DV name:	PROGPAR
DV label:	Participated in E/T program taken in last 12 months
DV value labels:	1 'Participated in E/T program'; 2 'Participated in E/T, but not in program'; 3 'Participated in E/T, but DK if in program'; 4 'Did not participate in any E/T'
Source question(s):	PROGDES2: Destination of E/T program taken in last 12 months (all respondents)
DV pseudo logic:	PROGDES2 is recoded into 4 categories specifying whether respondents participated in education and/or training and if they were in program or not.

Intensity of E/T program taken in last 12 months (students)

DV name:	PROGINT1
DV label:	Intensity of E/T program taken in last 12 months (students)
DV value labels:	1 'Full-time study in formal program'; 2 'Part-time study in formal program'; 3 'Equally full- and part- time study in formal program'; 4 'Studied in formal program, but DK intensity'; 5 'Studied in program, but not formal, intensity not known'; 6 'Studied in program, DK/R if formal, intensity not known'; 7 'Student but did not study in program'; 8 'Student but did not participate in any E/T'; 9 'Not a student at time of survey'
Source question(s):	D1: What is your current work situation? (Student is one of the choices.) F1: Did you take any education or training in last 12 months?

	F2: Did you take any courses as part of a program towards a certificate, diploma, or degree? F3: Was this program given by a school, college, or university? F6: Were you mostly a full-time or part-time student?
DV pseudo logic:	D1, F1, F2, F3, and F6 determine the intensity of education and/or training program taken in last 12 months. This variable is only for program participants who were students at time of survey.

Intensity of E/T program taken in last 12 months (all respondents)

DV name:	PROGINT2
DV label:	Intensity of E/T program taken in last 12 months (all respondents)
DV value labels:	1 'Full-time study in formal program'; 2 'Part-time study in formal program'; 3 'Equally full- and part- time study in formal program'; 4 'Studied in formal program but DK intensity'; 5 'Studied in program, but not formal, intensity not known'; 6 'Studied in program, DK/R if formal, intensity not known'; 7 'Did not study in program'; 8 'Did not participate in any E/T'
Source question(s):	F1: Did you take any education or training in last 12 months? F2: Did you take any courses as part of a program towards a certificate, diploma, or degree? F3: Was this program given by a school, college, or university? F6: Were you mostly a full-time or part-time student?
DV pseudo logic:	F1, F2, F3, and F6 determine the intensity of education and/or training program taken in last 12 months. This variable is for all respondents who reported participating in programs.

Duration of E/T program taken in last 12 months (hours)

DV name:	PROGDUR1
DV label:	Duration of E/T program taken in last 12 months (hours)
DV value labels:	Number of hours
Source question(s):	F7: Describe the total time spend in this program in terms of weeks, days or hours. F8: How many weeks did you spend in this program in last 12 months? F9: On average, how many hours per week was that? F10: In last 12 months, in total, how many days did you spend in this program? F11: On average, how many hours per day was that? F12: In last 12 months, how many hours in total did you spend in this program?
DV pseudo logic:	F7, F8, F9, F10, F11, F12, and PROGINT2 determine the number of hours spent in program of studies by all respondents. For those who responded to ‘Mostly full-time’ in F6 and were skipped over to question F13, the duration is imputed as 40 hours per week for 52 weeks.

Duration of E/T program taken in last 12 months (3 categories)

DV name:	PROGDUR2
DV label:	Duration of E/T program taken in last 12 months (3 categories)
DV value labels:	1 'Short duration - 40 hrs or less'; 2 'Medium duration - 41 to 100 hrs'; 3'Long duration - 101 hrs or more'
Source question(s):	PROGDUR1: Duration of E/T program taken in last 12 months (hours)

DV pseudo logic: PROGDUR1 is recoded into three categories.

Adult education population

DV name: ADEDPOP
DV label: Adult education population
DV value labels: 1 'adult education population'; 2 'non adult education population'
Source question(s): AGE_RESP: Age of the respondents.
PROGDES2: Program destination for all studying in last 12 months.
PROGINT2: Study intensity in last 12 months for all of those who studies as part of program.
F14B: Who contributed towards the direct expenses of this <program of studies>, that is, expense for tuition, course materials, travel, accommodation and so forth – is it an employer?
F14D: Who contributed towards the direct expenses of this <program of studies>, that is, expense for tuition, course materials, travel, accommodation and so forth – a union or professional association?
DV pseudo logic: F14B, F14D, AGE_RESP, PROGDES1 and PROGINT2 determine the population who participated in adult education and training. Full time students aged 16 to 25, in programs towards ISCED 0 – 6 and expense not contributed by an employer or a union or professional association, are excluded from the adult population.

Participation in any E/T taken in last 12 months

DV name: PELPART1
DV label: Participation in any E/T taken in last 12 months
DV value labels: 0 'Did not participate in any E/T'; 1 'Participated in E/T'
Source question(s): F1: Did you take any education or training in last 12 months?
DV pseudo logic: F1 is recoded into two categories distinguishing whether respondents participated in any education and/or training in last 12 months.

Participation in E/T program taken in last 12 months

DV name: PELPART2
DV label: Participation in E/T program taken in last 12 months
DV value labels: 0 'Did not participate in E/T program'; 1 'Participated in E/T program'
Source question(s): F2: Did you take any courses as part of a program towards a certificate, diploma, or degree?
DV pseudo logic: F2 is recoded into two categories distinguishing whether respondents participated in any education and/or training program in last 12 months.

Participation in one or more E/T courses taken in last 12 months

DV name: PELPART3
DV label: Participation in one or more E/T courses taken in last 12 months
DV value labels: 0 ‘Did not participate in E/T courses’; 1 'Participated in E/T courses’
Source question(s): F1: Did you take any education or training in last 12 months?
F15: Participated in any courses that were not part of your program.
F16: Participated in any courses that were not part of your program.
DV pseudo logic: F1, F15 and F16 determine whether respondents participated in one or more education and/or training courses that were not part of the program of studies in last 12 months.

DV name: PELPART4
DV label: Participation in informal types of learning undertaken in last 12 months
DV value labels: 0 'Did not participate in any informal types of learning'; 1 'Participated in informal types of learning'
Source question(s): F26A: Visit trade fairs, professional conferences or congresses.
F26B: Attend short lecture, seminars, workshops or special talks that were not part of a course.
F26C: Read manuals, reference books, journals or other written materials that were not part of a course.
F26D: Go on guided tours such as museums, art galleries, or other locations.
F26E: Use computers or the internet to learn but not as part of a course.
F26F: Use video, television, tapes to learn but not as part of a course.
F26G: Learn by watching, getting help from or advice from others but not from course instructors.
F26H: Learn by yourself by trying things out, doing things for practice, trying different approaches to doing things.
F26I: Learn by being sent around an organization to learn different aspects of that organization.
DV pseudo logic: F26A to F26I determine whether respondents participated in any informal types of learning in last 12 months. If they answered yes to at least on of these questions, they are categorized as participating in informal types of learning.

Participation in any E/T or learning in last 12 months (8 categories)

DV name: PELPART5
DV label: Participation in any E/T or learning in last 12 months (8 categories)
DV value labels: 1 'Participated in E/T program only'; 2 'Participated in E/T courses only'; 3 'Participated in informal types of learning only'; 4 'Participated in E/T program, courses & informal learning'; 5 'Participated in E/T program and E/T courses'; 6 'Participated in E/T program and informal types of learning'; 7 'Participated in E/T courses and informal types of learning'; 8 'Did not participate in any E/T or informal types of learning'
Source question(s): PELPART1: Participation in any E/T taken in last 12 months
PELPART2: Participation in E/T program taken in last 12 months
PELPART3: Participation in one or more E/T courses taken in last 12 months
PELPART4: Participation in informal types of learning undertaken in last 12 months
DV pseudo logic: PELPART1, PELPART2, PELPART3, and PELPART4 determine types of participation in education and/or training or informal types of learning in last 12 months.

Participation in any E/T taken in last 12 months (4 categories)

DV name: PELPART6
DV label: Participation in any E/T taken in last 12 months (4 categories)
DV value labels: 1 'Participated in E/T program only; 2 'Participated in E/T courses only'; 3 'Participated in E/T program, courses & informal learning'; 4 'Did not participate in any E/T or informal types of learning'

Source question(s): Same as PELPART5
DV pseudo logic: PELPART1, PELPART2, PELPART3, and PELPART4 determine types of participation in education and/or training or informal types of learning in last 12 months.

Number of courses taken in last 12 months that are not part of E/T program

DV name: ETCNUM
DV label: Number of courses taken in last 12 months that are not part of E/T program
DV value labels: 1 '1 course'; 2 '2 courses'; 3 '3 courses'; 4 '4 or more courses'
Source question(s): *For those who had taken courses as part of a program in last 12 months:*
F15: Did you participate in any courses that were not part of your program?
F15A: If yes, how many?
For those who had not taken any courses as part of a program in last 12 months:
F16: Did you participate in any courses that were not part of your program?
F16A: If yes, how many?
DV pseudo logic: F15, F15A, F16, F16A determine the number of courses taken in last 12 months that are not part of education and/or training program.

Duration of course 1 (hours)

DV name: ETCDUR1
DV label: Duration of course 1 (hours)
DV value labels: Number of hours
Source question(s): F18A: Would it be best describe the total time you spent in this course in terms of weeks, days, or hours?
F19A: In last 12 months, how many weeks did you spend in this course?
F20A: On average, how many hours per week was that?
F21A: In last 12 months, in total, how many days did you spend in this course?
F22A: On average, how many hours per day was that?
F23A: In last 12 months, how many hours in total was this course?
DV pseudo logic: F18A, F19A, F20A, F21A, F22A and F23A determine the duration of course 1 in hours.

Duration of course 2 (hours)

DV name: ETCDUR2
DV label: Duration of course 2 (hours)
DV value labels: Number of hours
Source question(s): F18B: Would it be best describe the total time you spent in this course in terms of weeks, days, or hours?
F19B: In last 12 months, how many weeks did you spend in this course?
F20B: On average, how many hours per week was that?
F21B: In last 12 months, in total, how many days did you spend in this course?
F22B: On average, how many hours per day was that?
F23B: In last 12 months, how many hours in total was this course?

DV pseudo logic: F18B, F19B, F20B, F21B, F22B and F23B determine the duration of course 2 in hours.

Duration of course 3 (hours)

DV name: ETCDUR3
DV label: Duration of course 3 (hours)
DV value labels: Number of hours
Source question(s): F18C: Would it be best describe the total time you spent in this course in terms of weeks, days, or hours?
F19C: In last 12 months, how many weeks did you spend in this course?
F20C: On average, how many hours per week was that?
F21C: In last 12 months, in total, how many days did you spend in this course?
F22C: On average, how many hours per day was that?
F23C: In last 12 months, how many hours in total was this course?
DV pseudo logic: F18C, F19C, F20C, F21C, F22C and F23C determine the duration of course 3 in hours.

Duration of all courses (hours)

DV name: ETCDUR1
DV label: Duration of all courses (hours)
DV value labels: Number of hours
Source question(s): ETCDUR1: Duration of course 1 (hours)
ETCDUR2: Duration of course 2 (hours)
ETCDUR3: Duration of course 3 (hours)
ETCNUM: Number of courses taken in last 12 months that are not part of E/T program
DV pseudo logic: ETCDUR1, ETCDUR2, ETCDUR3 and ETCNUM determine the total hours spent in all courses in last 12 months.

Household information and income

Have dependent children under age 16 living in the household

DV name: KIDSHOME
DV label: Have dependent children under age 16 living in the household
DV value labels: 0 ‘No children under 16 years old living in the household’; 1 ‘Children under 16 years old living in the household’
Source question(s): K1: Including yourself, how many people live in your household?
K2: Do you have any dependent children living with you in your household? (Children for whom you are financially and/or have sole or joint custody).
K3: What is the age of the youngest child in your household?
DV pseudo logic: K1, K2, and K3 determine whether there is/are dependent child/children under age of 16 living in the household.

Annual earnings from main job

DV name: EARNJOB1
DV label: Annual earnings from main job

DV value labels: Value of annual earning in national currency. *Canada (Canadian$), Italy (Euro), Bermuda (US$), and United States (US$), Norway (Norwegian Kroner), and Switzerland (Swiss Francs).*

Source question(s): D29: What <was/is> your status at this job or business?
D34: How many months in a year <did/do> you usually work at this job or business?
D35: <Did/do> you usually work every week of the month at this job or business?
D36: On average, how many weeks in a month <did/do> you usually work at this job or business?
D37: On average, how many hours per week <did/do> you usually work at this job or business?
D39: What is the easiest way for you to tell us your usual wage or salary for this job? Would it be hourly, weekly, annually, or on some other basis?
D40: How many days per week <did/do> you usually work at this job or business?
D41: What <was/is> your <fill text as indicated in D39, e.g. hourly, weekly, etc.> wage or salary before taxes and all other deductions at this job? Including tips and commissions?
D41A: Could only provide net earnings.
D42: Approximately how much <were/are> you paid per month at this job? Including tips and commissions and before taxes and deductions?
D42A: Could only provide net earnings.
D43: What <was/is> your annual personal net income before taxes and deductions from this business – that is, after all business expenses?

DV pseudo logic: D29, D34, D35, D36, D37, D39, D40, D41, D41A, D42, D42A and D43 are used to calculate the annual earnings from the main job. If respondent answered work(ed) every week of the month, 2*(365.25/(24*7)) is imputed as the number of weeks work(ed) at the job.

Monthly earnings from main job

DV name: EARNJOB2
DV label: Monthly earnings from main job
DV value labels: Value of monthly earnings in national currency. *Canada (Canadian$), Italy (Euro), Bermuda (US$), and United States (US$), Norway (Norwegian Kroner), and Switzerland (Swiss Francs).*
Source question(s): EARNJOB1: Annual earnings from main job
D34: How many months in a year <did/do> you usually work at this job or business?
DV pseudo logic: EARNJOB1 is divided by D34, namely the number of months work(ed) at main job.

Weekly earnings from main job

DV name: EARNJOB3
DV label: Weekly earnings from main job
DV value labels: Value of weekly earnings in national currency. *Canada (Canadian$), Italy (Euro), Bermuda (US$), and United States (US$), Norway (Norwegian Kroner), and Switzerland (Swiss Francs).*
Source question(s): EARNJOB1: Annual earnings from main job

D34: How many months in a year <did/do> you usually work at this job or business?
D36: On average, how many weeks in a month <did/do> you usually work at this job or business?

DV pseudo logic: EARNJOB1 is divided by D34 and D36, namely the number of months and weeks work(ed) at main job. If respondent answered work(ed) every week of the month, 2*(365.25/(24*7)) is imputed as the number of week work(ed) at main job.

Hourly earnings from main job

DV name: EARNJOB3
DV label: Weekly earnings from main job
DV value labels: Value of weekly earnings in national currency. *Canada (Canadian$), Italy (Euro), Bermuda (US$), and United States (US$), Norway (Norwegian Kroner), and Switzerland (Swiss Francs).*
Source question(s): EARNJOB1: Annual earnings from main job
D34: How many months in a year <did/do> you usually work at this job or business?
D36: On average, how many weeks in a month <did/do> you usually work at this job or business?
D37: On average, how many hours per week <did/do> you usually work at this job or business?
DV pseudo logic: EARNJOB1 is divided by D34, D36 and D37, namely the number of months, weeks and hours work(ed) at main job. If respondent answered work(ed) every week of the month, 2*(365.25/(24*7)) is imputed as the number of week work(ed) at main job.

Total personal income of respondent

DV name: INCTOT1
DV label: Total personal income of respondent
DV value labels: Value of personal income in national currency. *Canada (Canadian$), Italy (Euro), Bermuda (US$), and United States (US$), Norway (Norwegian Kroner), and Switzerland (Swiss Francs).*
Source question(s): K6: What is your best estimate of your personal income in <year before interview> from all sources, including those just mentioned, before taxes and deductions?
K7: Then, could you estimate whether your total personal income was <less than $20,000 or was it $20,000 or more>?
K8: Was it <less than $5,000>, <$5,000 to less than $10,000>, <$10,000 to less than $15,000>, <$15,000 or more>?
K9: Was it <less than $30,000>, <$30,000 to less than $40,000>, <$40,000 to less than $50,000>, <$50,000 to less than $60,000>, <$60,000 to less than $80,000>, <$80,000 or more>?
DV pseudo logic: Total person income is either reported directly from K6 or is estimated from the responses in K7, K8, or K9. If the income is imputed based on the intervals in K7, K8, or K9, the middle point of the interval is chosen. Users of this DV should be cautious about the last range, which for example is $80,000 or more in the case of Canada. The last interval was capped in order to arrive at middle point. In the case of Canada, a maximum of $99,999 was set.

Total household income of respondent

DV name: INCTOT2
DV label: Total household income of respondent
DV value labels: Value of personal income in national currency. *Canada (Canadian$), Italy (Euro), Bermuda (US$), and United States (US$), Norway (Norwegian Kroner), and Switzerland (Swiss Francs).*
Source question(s): K1: Including yourself, how many people live in your household?
K11: What is your best estimate of the total income of all household members (including yourself) from all sources in <year before interview> before taxes and deductions?
K12: Then, could you estimate whether the total income of all persons in your household was <less than $40,000 or was it $40,000 or more>?
K13: Was it <less than $10,000>, <$10,000 to less than $20,000>, <$20,000 to less than $30,000>, <$30,000 or more>?
K14: Was it <less than $50,000>, <$50,000 to less than $60,000>, <$60,000 to less than $80,000>, <$80,000 to less than $100,000>, <$100,000 or more>?
DV pseudo logic: If the respondent is the only person in the household, the total personal income INCTOT1 will be his/her total household income as well. Total household income is either reported directly from K11 or is estimated from the responses in K12, K13, or K14. If the income is imputed based on the ranges in K12, K13, or K14, the middle point of the range is chosen. Users of this DV should be cautious about the last range, which for example is $100,000 or more in the case of Canada. The last interval was capped in order to arrive at middle point. In the case of Canada, a maximum of $119,999 was set.

Total personal income of respondent in terms of purchasing power parity

DV name: INCPPP1
DV label: Total personal income of respondent in terms of purchasing power parity
DV value labels: Value of personal income in US$ (2001) PPP
Source question(s): INCTOT1: Total personal income of respondent
DV pseudo logic: INCTOT1 is divided by the purchasing power parity conversion factor. The purchasing power parity conversion factor is the number of units of a country's currency required to buy the same amount of goods and services in the domestic market as a US dollar would buy in the United States. The PPP conversion factors are from World Bank (latest figures from 2001). The PPP conversion factor for Bermuda is assumed to be 1.0 since this data was not available.

Total household income of respondent in terms of purchasing power parity

DV name: INCPPP2
DV label: Total household income of respondent in terms of purchasing power parity
DV value labels: Value of household income in US$ (2001) PPP
Source question(s): INCTOT2: Total household income of respondent
DV pseudo logic: INCTOT2 is divided by the purchasing power parity conversion factor. The purchasing power parity conversion factor is the number of units of a

country's currency required to buy the same amount of goods and services in the domestic market as a US dollar would buy in the United States. The PPP conversion factors are from World Bank (latest figures from 2001). The PPP conversion factor for Bermuda is assumed to be 1.0 since this data was not available.

Income support

DV name: INCSUPP
DV label: Income support
DV value labels: 0 'Did not receive any social income support'; 1 'Received social income support'
Source question(s): K5E: During <year prior to the interview>, did you receive any income from <Employment Insurance Benefits?>
K5F: During <year prior to the interview>, did you receive any income from <Social assistance and provincial supplements?>
DV pseudo logic: K5E and K5F determine whether respondent received social income support in the year before the survey.

Complex derived variables

Occupation and industry

Erikson and Goldthorpe’s class categories

DV name: EGP
DV label: Erikson and Goldthorpe’s class categories
DV scale: 1 'High skill service'; 2 'Low skill service'; 3 'Routine manual'; 4 'Self employed with employees'; 5 'Self employed without employees'; 7 'Manual supervisor'; 8 'Skilled manual'; 9 'Semi unskilled manual'; 10 'Farm labour'; 11 'Self employed farmer'
Source question(s): ISCOR: International Standardized Classification of Occupation (ISCO88) for respondent
DV pseudo logic: ISCO88 recoded to Erikson and Goldthorpe’s class categories.

International Socio-Economic Index – Respondent (continuous)

DV name: ISEI_R
DV label: International Socio-Economic Index – Respondent (continuous)
DV scale: Scores range from 16 to 90
Source question(s): ISCOR: Respondent’s occupation (International Standardized Classification of Occupation - ISCO88)
DV pseudo logic: ISCO88 recoded to International Socio-Economic Index.

International Socio-Economic Index – Respondent (quartiles)

DV name: ISEI_R_Q
DV label: International Socio-Economic Index – Respondent (quartiles)
DV scale: 1 ’Quartile 1’; 2 ’Quartile 2’; 3 ’Quartile 3’; ’Quartile 4’
Source question(s): ISEI_R: International Socio-Economic Index – Respondent (continuous)
DV pseudo logic: ISEI_R recoded into 4 categories.

Standard International Occupational Prestige Scale – Respondent (continuous)

DV name: SIOPS_R
DV label: Standard International Occupational Prestige Scale – Respondent (continuous)
DV scale: Scores range from 6 to 78
Source question(s): ISCOR: Respondent’s occupation (International Standardized Classification of Occupation - ISCO88)
DV pseudo logic: ISCO88 recoded to Standard International Occupational Prestige Scale.

Standard International Occupational Prestige Scale – Respondent (quartiles)

DV name: SIOPS_R_Q
DV label: Standard International Occupational Prestige Scale – Respondent (quartile)
DV scale: 1 ’Quartile 1’; 2 ’Quartile 2’; 3 ’Quartile 3’; ’Quartile 4’
Source question(s): SIOPS_R: Standard International Occupational Prestige Scale – Respondent (continuous)
DV pseudo logic: SIOPS_R recoded into 4 categories.

International Socio-Economic Index – Mother (continuous)

DV name: ISEI_M
DV label: International Socio-Economic Index – Mother (continuous)
DV scale: Scores range from 16 to 90
Source question(s): C3_C4: Mother's job when respondent was 16 years-old (International Standardized Classification of Occupation - ISCO88)
DV pseudo logic: ISCO88 recoded to International Socio-Economic Index.

International Socio-Economic Index – Mother (quartiles)

DV name: ISEI_M_Q
DV label: International Socio-Economic Index – Mother (quartiles)
DV scale: 1 ’Quartile 1’; 2 ’Quartile 2’; 3 ’Quartile 3’; ’Quartile 4’
Source question(s): ISEI_M: International Socio-Economic Index – Mother (continuous)
DV pseudo logic: ISEI_M recoded into 4 categories.

Standard International Occupational Prestige Scale – Mother (continuous)

DV name: SIOPS_M
DV label: Standard International Occupational Prestige Scale – Mother (continuous)
DV scale: Scores range from 6 to 78
Source question(s): C3_C4: Mother's job when respondent was 16 years-old (International Standardized Classification of Occupation - ISCO88)
DV pseudo logic: ISCO88 recoded to Standard International Occupational Prestige Scale.

Standard International Occupational Prestige Scale – Mother (quartiles)

DV name: SIOPS_M_Q
DV label: Standard International Occupational Prestige Scale – Mother (quartiles)
DV scale: 1 ’Quartile 1’; 2 ’Quartile 2’; 3 ’Quartile 3’; ’Quartile 4’
Source question(s): SIOPS_M: Standard International Occupational Prestige Scale – Mother (continuous)

DV pseudo logic: SIOPS_M recoded into 4 categories.

International Socio-Economic Index – Father (continuous)

DV name: ISEI_F
DV label: International Socio-Economic Index – Father (continuous)
DV scale: Scores range from 16 to 90
Source question(s): C7_C8: Father's job when respondent was 16 years-old (International Standardized Classification of Occupation - ISCO88)
DV pseudo logic: ISCO88 recoded to International Socio-Economic Index.

International Socio-Economic Index – Father (quartiles)

DV name: ISEI_F_Q
DV label: International Socio-Economic Index – Father (quartiles)
DV scale: 1 'Quartile 1'; 2 'Quartile 2'; 3 'Quartile 3'; 'Quartile 4'
Source question(s): ISEI_F: International Socio-Economic Index – Respondent (continuous)
DV pseudo logic: ISEI_F recoded into 4 categories.

Standard International Occupational Prestige Scale – Father (continuous)

DV name: SIOPS_F
DV label: Standard International Occupational Prestige Scale – Father (continuous)
DV scale: Scores range from 6 to 78
Source question(s): C7_C8: Father's job when respondent was 16 years-old (International Standardized Classification of Occupation - ISCO88)
DV pseudo logic: ISCO88 recoded to Standard International Occupational Prestige Scale.

Standard International Occupational Prestige Scale – Father (quartiles)

DV name: SIOPS_F_Q
DV label: Standard International Occupational Prestige Scale – Father (quartiles)
DV scale: 1 'Quartile 1'; 2 'Quartile 2'; 3 'Quartile 3'; 'Quartile 4'
Source question(s): SIOPS_F: Standard International Occupational Prestige Scale – Father (continuous)
DV pseudo logic: SIOPS_F recoded into 4 categories.

Knowledge and information class categories

DV name: OCCTYPE
DV label: Knowledge and information class categories
DV scale: 1 'Knowledge experts'; 2 'Managers'; 3 'Information high skill'; 4 'Information low skill'; 5 'Services low skill'; 6 'Manufacturing goods'
Source question(s): ISCOR: Respondent's occupation (International Standardized Classification of Occupation - ISCO88)
DV pseudo logic: ISCO88 recoded to knowledge and information class categories.

Knowledge-based classification of manufacturing and services

DV name: INDTYPE
DV label: Knowledge-based classification of manufacturing and services
DV scale: 1 'High-technology manufacturing'; 2 'Medium-high technology manufacturing'; 3 'Medium-low technology manufacturing'; 4 'Low

	technology manufacturing'; 5 'Knowledge intensive market services'; 6 'Public administration, defense, health'; 7 'Other community, social and personal services'; 8 'Utilities and construction'; 9 'Wholesale, retail, hotels and restaurants'; 10 'Transport and storage'; 11 'Primary industries'
Source question(s):	ISICR: Respondent's industry (International Standardized Industry Classification)
DV pseudo logic:	ISIC recoded to knowledge-based classification of manufacturing and services.

Literacy and numeracy practices at work

Index of reading engagement at work (IRT-continuous)

DV name:	WLIT1
DV label:	Index of reading engagement at work (IRT-continuous)
DV scale:	Standardized, with international mean and standard deviation set at (0, 1), adults aged 16 to 65 in following countries included in scaling: Bermuda, Canada, Italy, Norway, Switzerland, and United States.
Source question(s):	E1: How often <do/did> you read or use information from each of the following as part of your main job? Would you say at least once a week, less than once a week, rarely or never? E1A: Letters, memos or e-mails E1B: Reports, articles, magazines, journals E1C: Manuals, reference books including catalogues E1D: Diagrams or schematics E1E: Directions or instructions E1F: Bills, invoices, spreadsheets or budget tables
DV pseudo logic:	Confirmatory Factor Analysis (CFA) techniques are used to derive this variable. An international model is specified and tested in the Linear Structural Relations (LISREL) software package. Covariances and the accompanying asymptotic covariances of the source variables are estimated using the PRELIS software package and imported into LISREL. The Weighted Least Squares (WLS) method is used to estimate the model. The following fit statistics are used to assess the reliability and validity of the model: RMSEA, RMR, GFI, AGFI, NFI, NNFI, CFI and RFI. The international model meets the criteria of a Root Mean Square Error of Approximation (RMSEA) less than 0.08, Root Mean Square Residual (RMR) less than 0.05, and Adjusted Goodness of Fit Index (AGFI) greater than 0.90. In assessing fit, covariances among errors are allowed where theory supports the relationship. The final scaling of the DV is done using the Item Response Theory (IRT) technique in the PARSCALE software package.

Index of reading engagement at work (IRT-quartile)

DV name:	WLIT1_Q
DV label:	Index of reading engagement at work (quartiles)
DV value labels:	1 'Quartile 1'; 2 'Quartile 2'; 3 'Quartile 3'; 'Quartile 4'
Source question(s):	Same as WLIT1.
DV pseudo logic:	WLIT1 is recoded into 4 categories.

Index of reading engagement at work (sum scale)

DV name: E1SS
DV label: Index of reading engagement at work (sum scale)
DV value labels: 1 ’Never’; 2 ’Rarely’; 3 ’Less than once a week’; ’At least once a week’
Source question(s): See WLIT1.
DV pseudo logic: Source variables are recoded into ascending ordered category scales. The sum of values is divided by the number of source variables. Only cases with valid responses for all source variables are included.

Index of writing engagement at work (IRT-continuous)

DV name: WLIT2
DV label: Index of writing engagement at work (IRT-continuous)
DV scale: Same as WLIT1.
Source question(s): E2: How often <do/did> you write or fill out each of the following as part of your main job? Would you say at least once a week, less than once a week, rarely or never?
E2A: Letters, memos or e-mails
E2B: Reports, articles, magazines, journals
E2C: Manuals, reference books including catalogues
E2D: Directions or instructions
E2E: Bills, invoices, spreadsheets or budget tables
DV pseudo logic: Same as WLIT1.

Index of writing engagement at work (IRT-quartile)

DV name: WLIT2_Q
DV label: Index of writing engagement at work (quartiles)
DV value labels: 1 ’Quartile 1’; 2 ’Quartile 2’; 3 ’Quartile 3’; ’Quartile 4’
Source question(s): Same as WLIT2.
DV pseudo logic: WLIT2 is recoded into 4 categories.

Index of writing engagement at work (sum scale)

DV name: E2SS
DV label: Index of writing engagement at work (sum scale)
DV value labels: 1 ’Never’; 2 ’Rarely’; 3 ’Less than once a week’; ’At least once a week’
Source question(s): Same as WLIT2.
DV pseudo logic: Same as E1SS.

Index of numeracy engagement at work (IRT-continuous)

DV name: WLIT3
DV label: Index of numeracy engagement at work (IRT-continuous)
DV scale: Same as WLIT1.
Source question(s): E3: How often <do/did> you do each of the following as part of your main job? Would you say at least once a week, less than once a week, rarely or never?
E3B: Calculate prices, costs, or budgets
E3C: Count or read numbers to keep track of things
E3D: Manage time or prepare timetables
E3E: Give or follow directions or use maps or street directories
E3F: Use statistical data to reach conclusions

DV pseudo logic: Same as WLIT1.

Index of numeracy engagement at work (IRT-quartiles)

DV name: WLIT3_Q
DV label: Index of numeracy engagement at work (quartiles)
DV value labels: 1 'Quartile 1'; 2 'Quartile 2'; 3 'Quartile 3'; 'Quartile 4'
Source question(s): Same as WLIT3.
DV pseudo logic: WLIT3 is recoded into 4 categories.

Index of numeracy engagement at work (sum scale)

DV name: E3SS
DV label: Index of numeracy engagement at work (sum scale)
DV value labels: 1 'Never'; 2 'Rarely'; 3 'Less than once a week'; 4 'At least once a week'
Source question(s): E3: How often <do/did> you do each of the following as part of your main job? Would you say at least once a week, less than once a week, rarely or never?
E3A Measure or estimate the size or weight of objects
E3B: Calculate prices, costs, or budgets
E3C: Count or read numbers to keep track of things
E3D: Manage time or prepare timetables
E3E: Give or follow directions or use maps or street directories
E3F: Use statistical data to reach conclusions
DV pseudo logic: Same as E1SS.

Informal learning

Index of engaging in informal learning contexts (IRT-continuous)

DV name: INFL1
DV label: Index of engaging in informal learning contexts (IRT-continuous)
DV scale: Same as WLIT1.
Source question(s): F26: During the last 12 months, that is from <month and year> to <month and year> did you do any of the following learning activities?
F26A: Visit trade fairs, professional conferences or congresses
F26B: Attend short lecture, seminars, workshops or special talks that were NOT part of a course
F26D: Go on guided tours such as museums, art galleries, or other locations
F26I: Learn by being sent around an organization to learn different aspects of that organization
DV pseudo logic: Same as WLIT1.

Index of engaging in informal learning contexts (IRT-quartile)

DV name: INFL1_Q
DV label: Index of engaging in informal learning contexts (IRT-quartiles)
DV value labels: 1 'Quartile 1'; 2 'Quartile 2'; 3 'Quartile 3'; 'Quartile 4'
Source question(s): Same as INFL1.
DV pseudo logic: INFL1 is recoded into 4 categories.

Index of learning by doing (IRT-continuous)

DV name: INFL2
DV label: Index of engaging in learning by doing (IRT-continuous)
DV scale: Same as WLIT1.
Source question(s): F26: During the last 12 months, that is from <month and year> to <month and year> did you do any of the following learning activities?
- F26E: Use computers or the internet to learn but NOT as part of a course
- F26F: Use video, television, tapes to learn but NOT as part of a course
- F26G: Learn by watching, getting help from or advice from others but NOT from course instructors
- F26H: Learn by yourself by trying things out, doing things for practice, trying different approaches to doing things

DV pseudo logic: Same as WLIT1.

Index of learning by doing (IRT-quartiles)

DV name: INFL2_Q
DV label: Index of engaging in informal learning contexts (IRT-quartiles)
DV value labels: 1 'Quartile 1'; 2 'Quartile 2'; 3 'Quartile 3'; 'Quartile 4'
Source question(s): Same as INFL2.
DV pseudo logic: INFL2 is recoded into 4 categories.

Index of informal learning (sum scale)

DV name: F26SS
DV label: Index of informal learning (sum scale)
DV value labels: 1 'Yes'; 0 'No'
Source question(s): F26: During the last 12 months, that is from <month and year> to <month and year> did you do any of the following learning activities?
- F26A: Visit trade fairs, professional conferences or congresses
- F26B: Attend short lecture, seminars, workshops or special talks that were NOT part of a course
- F26C: Read manuals, reference books, journals or other written materials but not as part of a course
- F26D: Go on guided tours such as museums, art galleries, or other locations
- F26E: Use computers or the internet to learn but NOT as part of a course
- F26F: Use video, television, tapes to learn but NOT as part of a course
- F26G: Learn by watching, getting help from or advice from others but NOT from course instructors
- F26H: Learn by yourself by trying things out, doing things for practice, trying different approaches to doing things
- F26I: Learn by being sent around an organization to learn different aspects of that organization

DV pseudo logic: Same as E1SS.

Literacy and numeracy practices in daily life

Index of frequency and variety of reading (IRT-continuous)

DV name: HLIT1
DV label: Index of frequency and variety of reading (IRT-continuous)
DV scale: Same as WLIT1.
Source question(s): G3: How often do you read or use information from each of the following sources as part of your daily life? Please don't include time spent as part of your job or schooling. Would you say at least once a week, less than once a week, rarely or never?

- G3A: How often do you read or use information from newspaper?
- G3B: How often do you use or read information from magazines or articles
- G3C: How often do you read or use information from books fiction or non-fiction
- G3D: How often do you read or use information from letters, notes, e-mails

F26: During the last 12 months, that is from <month and year> to <month and year> did you do any of the following learning activities?

- F26C: Read manuals, reference books, journals or other written materials not as part of a course

DV pseudo logic: Same as WLIT1.

Index of frequency and variety of reading (IRT-quartiles)

DV name: HLIT1_Q
DV label: Index of frequency and variety of reading (IRT-quartiles)
DV value labels: 1 'Quartile 1'; 2 'Quartile 2'; 3 'Quartile 3'; 'Quartile 4'
Source question(s): Same as HLIT1.
DV pseudo logic: HLIT1 is recoded into 4 categories.

Index of frequency and variety of reading (sum scale)

DV name: G3SS
DV label: Index of frequency and variety of reading (sum scale)
DV value labels: 1 'Never'; 2 'Rarely'; 3 'Less than once a week'; 4 'At least once a week'
Source question(s): G3: How often do you read or use information from each of the following sources as part of your daily life? Please don't include time spent as part of your job or schooling. Would you say at least once a week, less than once a week, rarely or never?

- G3A: How often do you read or use information from newspaper?
- G3B: How often do you use or read information from magazines or articles
- G3C: How often do you read or use information from books fiction or non-fiction
- G3D: How often do you read or use information from letters, notes, e-mails

DV pseudo logic: Same as E1SS.

Index of attitude toward reading (IRT continuous)

DV name: HLIT4
DV label: Index of attitude toward reading (IRT-continuous)
DV scale: Same as WLIT1.
Source question(s): G7: Please tell me whether you strongly agree, agree, disagree, or strongly disagree with the following statements:
- G7C: I read only when I have to
- G7D: Reading is one of my favourite activities
- G7E: I enjoy talking about what I have read with other people

DV pseudo logic: Same as WLIT1.

Index of attitude toward reading (IRT quartiles)

DV name: HLIT4_Q
DV label: Index of attitude toward reading (IRT-quartiles)
DV value labels: 1 'Quartile 1'; 2 'Quartile 2'; 3 'Quartile 3'; 'Quartile 4'
Source question(s): Same as HLIT4.
DV pseudo logic: HLIT4 is recoded into 4 categories.

Index of attitude toward reading and numbers (sum scale)

DV name: G7SS
DV label: Index of attitude toward reading (sum scale)
DV value labels: 1 'Strongly disagree'; 2 'Disagree'; 3 'Agree'; 4 'Strongly agree'
Source question(s): G7: Please tell me whether you strongly agree, agree, disagree, or strongly disagree with the following statements:
- G7A: I am good with numbers and calculations
- G7B: I feel anxious when figuring such amounts as discounts, sales tax, or tips
- G7C: I read only when I have to
- G7D: Reading is one of my favourite activities
- G7E: I enjoy talking about what I have read with other people

DV pseudo logic: Same as E1SS. Note that when necessary, the substance associated with source variables is recoded into positively ordered categories (i.e., more is better than less).

Index of comfort with and perceived ability in math (IRT continuous)

DV name: HLIT5
DV label: Index of comfort with and perceived ability in math (IRT-continuous)
DV scale: Same as WLIT1.
Source question(s): G7: Please tell me whether you strongly agree, agree, disagree, or strongly disagree with the following statements:
- G7A: I am good with numbers and calculations

A9: Think about learning math and how you were taught math while a student at <insert ISCED level 2/3, lower or upper secondary level education>. Please tell me whether you strongly agree, agree, disagree, or strongly disagree with the following statements:
- A9A: I enjoy math in school
- A9B: I got good grades in math
- A9C: The teachers went too fast and I often got lost

A9D: I usually understood what was going on in math classes

DV pseudo logic: Same as WLIT1.

Index of comfort with and perceived ability in math (IRT quartiles)

DV name: HLIT5_Q
DV label: Index of comfort with and perceived ability in math (IRT-quartiles)
DV value labels: 1 'Quartile 1'; 2 'Quartile 2'; 3 'Quartile 3'; 'Quartile 4'
Source question(s): Same as HLIT5.
DV pseudo logic: HLIT5 is recoded into 4 categories.

Index of comfort with and perceived ability in math (sum scale)

DV name: A9SS
DV label: Index of comfort with and perceived ability in math (sum scale)
DV value labels: 1 'Strongly disagree'; 2 'Disagree'; 3 'Agree'; 4 'Strongly agree'
Source question(s): A9: Think about learning math and how you were taught math while a student at <insert ISCED level 2/3, lower or upper secondary level education>. Please tell me whether you strongly agree, agree, disagree, or strongly disagree with the following statements:

A9A: I enjoy math in school
A9B: I got good grades in math
A9C: The teachers went too fast and I often got lost
A9D: I usually understood what was going on in math classes

DV pseudo logic: Same as E1SS. Note that when necessary, the substance associated with source variables is recoded into positively ordered categories (i.e., more is better than less).

Index of visiting library/bookstore (sum scale)

DV name: G1SS
DV label: Index of visiting library and bookstore (sum scale)
DV value labels: 1 'Never'; 2 'Once or twice during the year'; 3 'Several times during the year'; 4 'Monthly'; 5 'Weekly'
Source question(s): G1: About how often do you do the following activities whether these activities are done in person or on computer? Would that be weekly, monthly, several times during the year, once or twice during the year, or never?

G1A: Use a library
G1B: Visit a bookstore

DV pseudo logic: Same as E1SS.

Index of newspaper reading variety (sum scale)

DV name: G5SS
DV label: Index of newspaper reading variety (sum scale)
DV value labels: 1 'Yes'; 0 'No'
Source question(s): G5: I am now going to read you a list of some different parts of a newspaper. Please tell me which parts you generally read when looking at a newspaper.

G5A: National/international news
G5B: Regional or local news
G5C: Sports

G5D: Home, fashion, food or health
G5E: Editorial page
G5F: Financial news or stock listings
G5G: Book, movie or art review
G5H: Advice column

DV pseudo logic: Same as E1SS.

Information communications technologies

Index of perceived usefulness and attitude toward computers (IRT-continuous)

DV name: ICT1
DV label: Index of perceived usefulness and attitude toward computers (IRT-continuous)
DV scale: Same as WLIT1.
Source question(s): H15: Please tell me whether you strongly agree, agree, disagree, or strongly disagree with each of the following statement:

- H15C: Computers have made it possible for me to get more done in less time
- H15D: Computers have made it easier for me to get useful information
- H15E: Computers have helped me to learn new skills other than computer skills
- H15F: Computers have helped me to communicate with people
- H15G: Computers have helped me reach my occupational (career) goals

DV pseudo logic: Same as WLIT1.

Index of perceived usefulness and attitude toward computer (IRT-quartiles)

DV name: ICT1_Q
DV label: Index of perceived usefulness and attitude toward computers (IRT-quartiles)
DV value labels: 1 'Quartile 1'; 2 'Quartile 2'; 3 'Quartile 3'; 'Quartile 4'
Source question(s): Same as ICT1.
DV pseudo logic: ICT1 is recoded into 4 categories.

Index of perceived usefulness and attitude toward computers (sum scale)

DV name: H15SS
DV label: Index of perceived usefulness and attitude toward computer (sum scale)
DV value labels: 1 'Strongly disagree'; 2 'Disagree'; 3 'Agree'; 4 'Strongly agree'
Source question(s): H15: Please tell me whether you strongly agree, agree, disagree, or strongly disagree with each of the following statement:

- H15A: My level of computer skills meets my present needs
- H15B: I feel comfortable installing or upgrading computer software
- H15C: Computers have made it possible for me to get more done in less time
- H15D: Computers have made it easier for me to get useful information

H15E: Computers have helped me to learn new skills other than computer skills
H15F: Computers have helped me to communicate with people
H15G: Computers have helped me reach my occupational (career) goals

DV pseudo logic: Same as E1SS.

Index of diversity and intensity of internet use (IRT-continuous)

DV name: ICT3
DV label: Index of diversity and intensity of internet use (IRT-continuous)
DV scale: Same as WLIT1.
Source question(s): H5: In a typical month, how often did you use the Internet for the following purposes? (Daily, a few times a week, a few times a month, never).

H5A: Electronic mail (email)
H5B: Participate in chat groups or other on-line discussions
H5C: Shopping (including browsing for products or services but not necessarily buying)
H5D: Banking
H5E: Formal education or training (part of a formal learning activity such as a course or a program of studies)
H5F: Obtain or save music
H5G: Read about news and current events
H5H: Search for employment opportunities
H5I: Search for health related information
H5J: Search for weather related information
H5K: Search for government information
H5L: Playing games with others
H5M: General browsing
H5N: Other purposes; specify:

H13: In a typical month, how many hours did you use a computer at home

DV pseudo logic: Same as WLIT1.

Index of diversity and intensity of internet use (IRT-quartiles)

DV name: ICT3_Q
DV label: Index of diversity and intensity of internet use (IRT-quartiles)
DV value labels: 1 'Quartile 1'; 2 'Quartile 2'; 3 'Quartile 3'; 'Quartile 4'
Source question(s): Same as ICT3.
DV pseudo logic: ICT3 is recoded into 4 categories.

Index of diversity and intensity of internet use (sum scale)

DV name: H5SS
DV label: Index of diversity and intensity of internet use (sum scale)
DV value labels: 1 'Never'; 2 'A few times a month'; 3 'A few times a week'; 4 'Daily'
Source question(s): H5: In a typical month, how often did you use the Internet for the following purposes? (Daily, a few times a week, a few times a month, never).

H5A: Electronic mail (email)

H5B: Participate in chat groups or other on-line discussions
H5C: Shopping (including browsing for products or services but not necessarily buying)
H5D: Banking
H5E: Formal education or training (part of a formal learning activity such as a course or a program of studies)
H5F: Obtain or save music
H5G: Read about news and current events
H5H: Search for employment opportunities
H5I: Search for health related information
H5J: Search for weather related information
H5K: Search for government information
H5L: Playing games with others
H5M: General browsing
H5N: Other purposes; specify:

DV pseudo logic: Same as E1SS.

Index of using computers for task oriented purposes (IRT-continuous)

DV name: ICT6
DV label: Index of using computers for task oriented purposes (IRT-continuous)
DV scale: Same as WLIT1.
Source question(s): H3: In a typical month, how often did you use a computer for the following purposes? (Daily, a few times a week, a few times a month, never).

H3A: Writing or editing text
H3B: Accounts, spreadsheets or statistical analysis
H3C: Creating graphics, designs, pictures, or presentation
H3D: Programming or writing computer code
H3E: Keeping a schedule or calendar
H3F: Reading information on a CD-ROM or DVD
H13: In a typical month, how many hours did you use a computer at home

DV pseudo logic: Same as WLIT1.

Index of using computers for task oriented purposes (IRT-quartiles)

DV name: ICT6_Q
DV label: Index of using computers for task oriented purposes (IRT-quartiles)
DV value labels: 1 'Quartile 1'; 2 'Quartile 2'; 3 'Quartile 3'; 'Quartile 4'
Source question(s): Same as ICT6.
DV pseudo logic: ICT6 is recoded into 4 categories.

Index of using computers for task oriented purposes (sum scale)

DV name: H3SS
DV label: Index of using computers for task oriented purposes (sum scale)
DV value labels: 1 'Never'; 2 'A few times a month'; 3 'A few times a week'; 4 'Daily'
Source question(s): H3: In a typical month, how often did you use a computer for the following purposes? (Daily, a few times a week, a few times a month, never).

H3A: Writing or editing text
H3B: Accounts, spreadsheets or statistical analysis

H3C: Creating graphics, designs, pictures, or presentation
H3D: Programming or writing computer code
H3E: Keeping a schedule or calendar
H3F: Reading information on a CD-ROM or DVD
H3G: Playing games
H3H: Other purposes; specify:

DV pseudo logic: Same as E1SS.

Index of intensity and variety of technology use in general (sum scale)

DV name: H1SS
DV label: Index of intensity and variety of technology use in general (sum scale)
DV value labels: 1 'Never'; 2 'A few times a month'; 3 'A few times a week'; 4 'Daily'
Source question(s): H1:In a typical month, how often did you use each of the following?

H1A: A cellular phone
H1B: Calculator
H1C: A fax machine (i.e., a stand alone fax machine, not fax software on a computer.)
H1D: A touch-tone phone to buy something, pay a bill, financial transactions
H1E: An automated banking machine (e.g., ATM, Bankomat…)

DV pseudo logic: Same as E1SS.

Index of intensity and variety of technology use at work (sum scale)

DV name: H10SS
DV label: Index of intensity and variety of technology use at work (sum scale)
DV value labels: 1 'Yes'; 0 'No'
Source question(s): H10: In a typical month, did you use any of the following technologies in your job?

H10A: Calculator
H10B: Cell phone
H10C: Fax machine
H10D: Electronic personal organizer

DV pseudo logic: Same as E1SS.

Index of using computers in multiple locations (sum scale)

DV name: H14SS
DV label: Index of using computers in multiple locations (sum scale)
DV value labels: 1 'Yes'; 0 'No'
Source question(s): H14: In a typical month, did you use a computer at…

H14A: A friend's home
H14B: A relative's home
H14C: A public library (excluding the library catalogue)
H14D: An internet café
H14E: A community resource centre (e.g., employment centre)
H14F: A school, educational or training institution
H14G: Some other location, specify:

DV pseudo logic: Same as E1SS.

Civic participation

Index of engaging in community groups/organizations (IRT-continuous)

DV name: CIVI1
DV label: Index of engaging in community groups or organizations (IRT-continuous)
DV scale: Same as WLIT1.
Source question(s): G8: The next questions are about your volunteer work and the organizations in which you participate. During the last 12 months, did you participate in any of the following groups or organizations?

- G8A: A political organization
- G8B: A sports or recreation organization (e.g. Baseball League, Tennis Club, etc.)
- G8C: A cultural, education or hobby group (e.g. theatre Group, Book Club, Bridge Club, etc.)
- G8D: A service club (e.g. Kiwanis, Knights of Columbus, Shriners)
- G8E: A neighbourhood, civic or community association or a school group (e.g. Parent/Teachers Association, your neighbourhood community association)

DV pseudo logic: Same as WLIT1.

Index of engaging in community groups/organizations (IRT-quartiles)

DV name: CIVI1 _Q
DV label: Index of engaging in community groups or organizations (IRT-quartiles)
DV value labels: 1 'Quartile 1'; 2 'Quartile 2'; 3 'Quartile 3'; 'Quartile 4'
Source question(s): Same as CIVI1.
DV pseudo logic: CIVI1 is recoded into 4 categories.

Index of engaging in community groups/organizations (sum scale)

DV name: G8SS
DV label: Index of engaging in community groups or organizations (sum scale)
DV value labels: 1 'Yes'; 0 'No'
Source question(s): G8: The next questions are about your volunteer work and the organizations in which you participate. During the last 12 months, did you participate in any of the following groups or organizations?

- G8A: A political organization
- G8B: A sports or recreation organization (e.g. Baseball League, Tennis Club, etc.)
- G8C: A cultural, education or hobby group (e.g. theatre Group, Book Club, Bridge Club, etc.)
- G8D: A service club (e.g. Kiwanis, Knights of Columbus, Shriners)
- G8E: A neighbourhood, civic or community association or a school group (e.g. Parent/Teachers Association, your neighbourhood community association)
- G8F: A group associated with a community of worship (e.g. a youth group associated with a church)
- G8G: Any other group or organization

DV pseudo logic: Same as E1SS.

Index of engaging in community groups/organizations as unpaid volunteer (IRT-continuous)

DV name: CIVI2
DV label: Index of engaging in community groups or organizations as an unpaid volunteer (IRT-continuous)
DV scale: Same as WLIT1.
Source question(s): G9: In the last 12 months, did you do any of the following activities as an UNPAID VOLUNTEER through a group or organization?
G9A: Fundraising
G9B: Serving as an unpaid member of a board
G9C: Coaching, teaching or counselling
G9D: Collecting food or other goods for charity
G9E: Any other activities such as (organizing/supervising events; office work or providing information on behalf of an organization)
DV pseudo logic: Same as WLIT1.

Index of engaging in community groups/organizations as unpaid volunteer (IRT-quartiles)

DV name: CIVI2 _Q
DV label: Index of engaging in community groups or organizations as an unpaid volunteer (IRT-quartiles)
DV value labels: 1 'Quartile 1'; 2 'Quartile 2'; 3 'Quartile 3'; 'Quartile 4'
Source question(s): Same as CIVI2.
DV pseudo logic: CIVI2 is recoded into 4 categories.

Index of engaging in community groups/organizations as unpaid volunteer (sum scale)

DV name: G9SS
DV label: Index of engaging in community groups or organizations as an unpaid volunteer (sum scale)
DV value labels: 1 'Yes'; 0 'No'
Source question(s): Same as CIVI2.
DV pseudo logic: Same as E1SS.

Health and well being

Index of general satisfaction and feeling good (IRT-continuous)

DV name: WELL1
DV label: Index of general satisfaction and feeling good (IRT-continuous)
DV value labels: Same as WLIT1.
Source question(s): G10: On the whole, how do you feel about your life over the past 12 months? Would you say that you are extremely satisfied, satisfied, unsatisfied or extremely unsatisfied?
G16: These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give the one answer that comes closest to the way you have been feeling. (All

of the time, most of the time, a good bit of the time, some of the time, a little of the time, none of the time). How much of the time during the past 4 weeks…
G16A: Have you felt calm and peaceful? Would that be…
G16B: Did you have a lot of energy? Would that be…
G16C: Have you felt downhearted and blue? Would that be…

DV pseudo logic: Same as WLIT1.

Index of general satisfaction and feeling good (IRT-quartiles)

DV name:	WELL1_Q
DV label:	Index of general satisfaction and feeling good (IRT-quartiles)
DV value labels:	1 'Quartile 1'; 2 'Quartile 2'; 3 'Quartile 3'; 'Quartile 4'
Source question(s):	Same as WELL1.
DV pseudo logic:	WELL1 is recoded into 4 categories.

Index of interference of physical health with work activities (IRT-continuous)

DV name:	WELL2
DV label:	Index of interference of physical health with work activities (IRT-continuous)
DV value labels:	Same as WLIT1.
Source question(s):	G13: During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your physical health? G13A: Accomplished less than you would like G13B: Were limited in the kind of work or other activities G15: During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)? Was this…(Not at all, a little bit, moderately, quite a bit, extremely).
DV pseudo logic:	Same as WLIT1.

Index of interference of physical health with work activities (IRT-quartiles)

DV name:	WELL2_Q
DV label:	Index of interference of physical health with work activities (IRT-quartiles)
DV value labels:	1 'Quartile 1'; 2 'Quartile 2'; 3 'Quartile 3'; 'Quartile 4'
Source question(s):	Same as WELL2.
DV pseudo logic:	WELL2 is recoded into 4 categories.

Physical component summary -- SF 12 health index (continuous)

DV name:	PCS
DV label:	Physical component summary -- SF 12 health index (continuous)
DV value labels:	The scale for PCS is from 9.94738 to 70.13284.
Source question(s):	G11: In general, would you say your health is excellent, very good, good, fair, poor? G12: The following questions are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much – would that be limited a lot or limited a little? G12A: Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling or playing golf G12B: Climbing several flights of stairs

	G13: During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your physical health? G13A: Accomplished less than you would like G13B: Were limited in the kind of work or other activities G14: During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)? G14A: Accomplished less than what you like G14B: Didn't do work or other activities as carefully as usual G15: During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)? Was this…(Not at all, a little bit, moderately, quite a bit, extremely). G16: These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give the one answer that comes closest to the way you have been feeling. (All of the time, most of the time, a good bit of the time, some of the time, a little of the time, none of the time). How much of the time during the past 4 weeks… G16A: Have you felt calm and peaceful? Would that be… G16B: Did you have a lot of energy? Would that be… G16C: Have you felt downhearted and blue? Would that be… G17: During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting friends, relatives, etc.)? Was it all of the time, most of the time, some of the time, a little of the time, none of the time.
DV pseudo logic:	In the SF-12 health survey version 1, the constant (1990) for physical weight is 56.57706. Each variable features one or more categories, where each category is assigned a physical weight; the value could be either positive or negative. The physical weights associated with the category chosen in each variable are cumulatively added to the constant (1990) to compute the final PCS scale.

Physical component summary -- SF 12 health index (quartiles)

DV name:	PCS_Q
DV label:	Physical component summary -- SF 12 health index (quartile)
DV value labels:	1 'Quartile 1'; 2 'Quartile 2'; 3 'Quartile 3'; 'Quartile 4'
Source question(s):	PCS: Physical component summary -- SF 12 health index (continuous)
DV pseudo logic:	PCS is recoded into 4 categories.

Mental component summary -- SF 12 health index (continuous)

DV name:	MCS
DV label:	Mental component summary -- SF 12 health index (continuous)
DV value labels:	The scale for MCS is from 5.89058 to 72.28379.
Source question(s):	Same as PCS.
DV pseudo logic:	In the SF-12 health survey version 1, the constant (1990) for physical weight is 60.75781. Each variable features one or more categories, where each category is assigned a physical weight; the value could be either positive or negative. The physical weights associated with the category

chosen in each variable are cumulatively added to the constant (1990) to compute the final MCS scale.

Mental component summary -- SF 12 health index (quartiles)

DV name: MCS_Q
DV label: Mental component summary -- SF 12 health index (quartile)
DV value labels: 1 'Quartile 1'; 2 'Quartile 2'; 3 'Quartile 3'; 4 'Quartile 4'
Source question(s): MCS: Mental component summary -- SF 12 health index (continuous)
DV pseudo logic: MCS is recoded into 4 categories.

Self Assessment

Index of satisfaction of own skills for work (IRT-continuous)

DV name: SELF1
DV label: Index of satisfaction of own skills for work (IRT-continuous)
DV value labels: Same as WLIT1.
Source question(s): E4: Please tell me whether you strongly agree, agree, disagree, or strongly disagree with the following statements:
E4A: I have reading skills in <insert language> I need to do my main job well
E4B: I have writing skills in <insert language> I need to do my main job well
E4C: I have the math skills I need to do my main job well
DV pseudo logic: Same as WLIT1.

Index of satisfaction of own skills for work (IRT-quartile)

DV name: SELF1_Q
DV label: Index of satisfaction of own skills for work (IRT-quartile)
DV value labels: 1 'Quartile 1'; 2 'Quartile 2'; 3 'Quartile 3'; 4 'Quartile 4'
Source question(s): SELF1: Index of satisfaction of own skills for work (IRT-continuous)
DV pseudo logic: SELF1 is recoded into 4 categories.

Index of satisfaction of own skills for work (sum scale)

DV name: E4SS
DV label: Index of satisfaction of own skills for work (sum scale)
DV value labels: 1 'Strongly disagree'; 2 'Disagree'; 3 'Agree'; 4 'Strongly agree'
Source question(s): Same as SELF1.
DV pseudo logic: Same as E1SS.

6.4.7 Creation of Master file (Step 7)

The Master file is a final clean file with all the necessary variables, from which the PUMF and International Microdata files can be created. To create this file, the data had to go through a series of functions, like recodes, renames, linkages, etc. Many issues that were not resolved in the earlier steps, due of checks from frequency counts, missing values, outliers, were dealt in this step. This step was performed right after the Imputation step.

Rename variables

In order to be compatible with the international standards, the variable names for the complete questionnaire were renamed. In each variable the 'Q' at the beginning was removed (example: QA1 became A1). No change was done to the variables that were either created as per International requirements or had no bearing on the international file. Some variables that were appended to the file from different sources were renamed due to the variable name length being too long, for example imp_tqlang2 to itlang2, fcorestat to fcorest, fmainstat to fmainst, finalstat to finalst, taskstatus to TaskStat, MotherTongue to MotherTo, Immigrant to Immigrnt, Aboriginal to Aborig, to name a few (normal length being 8-bytes long). If a variable was created to indicate a flag or an imputation, then the naming convention was created accordingly. The flag variables like BQStatus was changed to BQStFlag and imp_age, an imputation variable, was changed to iageflag for example, yet still others were renamed to follow collection standards like BQLP_Q01 to LP_Q01 and Primary to SampleID.

New variables

Several new variables were created to add information to the exiting file and also to accommodate information required for the International file. Some were created from the exiting variables, and some from merging different files. Methodology created many new variables, besides final status, that were required for both National and International file. They are detailed below but it should be pointed out that the majority of the variables added at this step were administrative in nature and are not part of the Public Use Microdata file.

Merging data sets

The Survey team created variables like Rural/Urban (Urban and Rural designations), Stratum (geographic indicator), DA (Dissemination Area), that were also required for both files, National and International. It also imputed missing values for variables for the National file like Province, Mother tongue, Immigrant status (Born in Canada), Aboriginal status. These new variables were merged to our existing Main file and hence added to variable list.

Merging the IRLFLAGS dataset

Several variables were created for Methodological reasons and kept in a file called IRLFLAGS (short for International Record Layout Flags). For instance, given the detailed geographic information available Rural/Urban areas were defined and a variable created based on Census information regarding the appropriate community size for each respondent household.

Definition for the UARAtype variable, was directly taken from census, whereas RUFLAG variable was recreated, to condense the subdivision with respect to population density. Final statuses were also created for the BQ, Core and Main, using information from combination of variables. To create Final statuses for both Core and Main, methodology used a combination of 2 two files, namely core1.txt (variable=CS_Q02) and out_core.txt (variable=CSTAT) for core and main.txt (variable=MT_Q02) and out_main.txt (variable=MSTAT) for main.

Methodology determined the final BQ status based on complete, partial and incomplete responses to each module.

Complete BQ data (23,098 records): At least 75% of the modules were at least 75% complete, plus gender, age and education were reported.

Partial BQ data (179 records): At least 50% of the modules were at least 75% complete plus gender, age and education were reported.

Incomplete BQ data (472 records): Less than 50% of the modules were 75% complete or one of gender, age or education (total years of schooling or highest level achieved) were missing.

BQ status was established by coding 2 for a complete record, 1 for a partially complete record and 0 for an incomplete record based on the BQ data
The number of respondents in the response file before this step was 23, 749. A total of 711 records were dropped: 472 records with incomplete BQ data (as defined above), 169 records with Partial BQ data and no task booklet, and 70 records with suspect data from possibly contaminated responses. The total number of respondents after this step was 23, 038 respondents.

Merging the DV dataset

Some questions were asked in the Entry component of the survey and some in the Background component. When the respondent and Entry person (Screener) were the same, then the question was not repeated, otherwise, it was asked again. The survey team combined the two responses from the divergent files into a common question that contained one response per record. The DV (Derived Variable) dataset therefore, helped us to merge Mother Tongue, Immigration status and Aboriginal status. This line of questioning was developed to reduce respondent burden.

Merging the TQLANG2 dataset

TQLANG2 (Country ID) was created according to the linguistic profiles of each respondent and to be later used for the International file as CNTRID. The information for TQLANG2 was derived from TQLANG (Language of Task Booklet). When it was missing, the survey team imputed these values by using a donor record, with a similar linguistic profile. A flag was also created called ITQLANG2 to indicate, when the record was imputed and when it was not.

Merging the DA file

In order to facilitate the creation of small area literacy estimates, the Dissemination Area (DA) of the household was appended by the survey team according to structures defined by Census.

Merging the Weights data set

The survey team also created a series of weights required to assess the quality of the survey and to allow users the ability to accurately calculate the standard error of proficiency estimates. The variables appended at this stage include:

THEORWT: The Theoretical Weight corresponds to the inverse of the probability of selection. These are provided by the sample design. This weight was used for quality analysis.

SUBWT: The Sub-Weight is the product of THEORWT by the non-response adjustment factor.

POPWT: Population weight. This variable should be used to allow the sample up to the ability to represent the Canadian population 16 years and older. It is the product of SUBWT by the adjustment to known totals factor. This is the individual's overall sampling weight. The sum of these weights within a sample adds up to the size of the population.

REPLIC01 to REPLIC30: This is a series of 30 replicate weights, that, when used in conjunction with the 5 plausible ability values (prose, document, numeracy, health or problem solving), can produce accurate population point estimates with calculable standard errors. The procedures for the proper use of these variables can be found in Chapter 8 (Data Quality).

The Population weight and all replicate weights are included on the Public Use Microdata file.

Merging the Casevnt file

New variables were also created by the processing unit from either a variable itself or from using a combination of variables. Casevnt file was merged in order to create several variables like C01 (total number of visits to the household by the interviewer) and C07_1 to C07_10 (status of these visits, like no answer, refused, respondent has moved or respondent interviewed). The values for C07_1 to C07_10 were derived using some values of APLCSTAT, whereas, BrnInCan was created using two variables, namely, A1 and Imm_Q01.

Merging the Times file

End time and start time for different components of the Survey were also generated from theTimes file returned by the CAPI application. These included the start and end dates and times for each survey component (Background Questionnaire, Core Task Booklet and Main Task Booklet. The "Times" file has several records for the same household, similar to the CASEVNT file, therefore, when an identical match occurred, the required information from the record was merged to the main file.

Merging the COREFIN file

Variables like CS_Q02 (Detail of physical and emotional conditions), SCORE (total calculated score from doing the Core tasks), C_QA1 (language of Core task) and OUTCORE (core status at the end of core task) were captured. Some of the information from this file was merged to create new variables, while other raw variables were imported as is in order to add more information to the existing file.

Merging the MAINFIN file

Variables like MT_R29 (calculator usage), MS_Q02 (Detail of physical and emotional conditions at the time of doing the Main Task Booklet) and OUTMAIN (main status at the end of Task Booklet) were merged to the main file. Some of these variables from this file were used to create new variables like CBLOCK5, CBLOCK6, CBLOCK56 from MT_R29, while others were just to add more information to the existing file.

Merging the Bookflag dataset

When checking MainTask Booklets for valid and invalid answers, it was noticed that there were marks in some Booklets, which could have been marked by the respondents or the interviewer may have erroneously instructed the respondent to make such marks to display that they'd seen the information and attempted the question. A file was created to identify each such case, by creating a flag of 1 (mark in booklets) or 2 (interviewers may have been instructed). This file was merged to the Main file to add 'BookFlag' variable

Merging the CanadaScore, B8BQ3D files and corescore dataset

Separate files were created for core scores (COREQ1S1 to COREQ6S2), where each question was given a score of 1, if answered correctly and 0 if not correct. Each question was checked twice, therefore 2 scores were given per question, score1 for the 1st score (S1) and score 2 (S2) for the second check. There were a total of 6 questions. When the total of scores added to 3 or more, then Main Task booklet component was generated, allowing respondents to further answer more detailed questions on Literacy and Numeracy. A variable called "SCORE" was created in this file, to sum the scores of all the six questions. Two variables, B8BQ3DS1 (score 1) & B8BQ3DS2 (score 2) were created to accommodate an error found in one of the French items (please consult the quality notes for specific issues regarding the Canadian dataset.

Merging the BckGrn file

Original values of some variables were recoded in previous step. In order to re-capture the original values, the BckGrn file was merged, by keeping only the required variables. Such variables were QA1, BQIQ01 & BQLP_Q01. QA1 helped reload A1 ("Were you born in Canada" question), with original values. This question was also used to derive BrnInCan (Born in Canada) variable, using BQIQ01(from the BQ) and IMM_Q01(from the Entry module). BQLP_Q01 (language of interview) helped repopulate LP_Q01. BQIQ01 was also picked, to repopulate original values. However, when the values of this variable was missing they were made '6' (valid skip).

Merging the Entry file

Entry file was merged because a key variable (BOOKID) used to identify which of the 28 booklets was ultimately administered to the respondent was missing a value for 5 records. For this reason, the variable "BookLtID", a variable identifying which of the 28 booklets was assigned to each respondent on the sample file, was retrieved from the Entry file and the value of this variable was used to impute a value to BOOKID.

Merging the Health Literacy Scores

Following work done at the Educational Testing Services, the Prose, Document and Numeracy domains were found to present sufficient health related information to yield a fourth psychometric measure known as Health Literacy. In practical terms what this the tasks measure a persons' ability to use and understand text and documents containing health related information. In order to add value to the dataset, the five plausible values for this added domain were calculated for the Canadian IALSS data and were merged with the dataset in this step.

Deriving Education variables

Education variables for the international file were also derived using a combination of variables.

Thus, A4_ISCED, the highest level of education reached by the responden,t was derived using questions A4A, A4B, A4C and A4F. These received International Standard Classification of Education (ISCED) codes to facilitate international comparisons. Once verified, these codes could then be used to recreate the Canadian education classifications found in variable A4. The assured concordance between the two classification systems.

Similarly, question C2 was used to create C2_ISCED (education of respondent's Mother) and C6 was used to create C6_ISCED (education of respondent's father) Finally, F4_ISCED was created from F4, an adult education variable, that measured whether the respondent has been taking or has taken courses toward a further degree or diploma. As the name implies, this was coded to the ISCED system and , like A4, code back to the Canadian labels.

Assigning missing values

Missing values in variables were recoded to either '9'/'99' or '6'/'96' depending upon the logic of the question. Some missing values were derived by using values from another variable. A good example is, of BookID, when it was missing then values of BookLtID were picked up.

Recodes

Some variables were not compatible with the International file and therefore needed to be adjusted. A good example again was "don't know" value in questions A9A – A9D. In the international file this reply was not an option, for No opinion was considered basically same as don't know, therefore values of '7' (don't know) in Q9A- Q9D were recoded to '5' (no opinion). Value of '7' in A10 was changed to '8' in order to meet international standard.

Recoding B1A, B1B and B1C

Question B1, measured the "language first learned and still understood" (Mother tongue). However, in order to reflect the Canadian multilingual reality, the IALSS had room for up to 3 choices provided they were all learned and understood at the same time (as subjectively interpreted by the respondents). These three choices needed to be distilled into one useable variable. This variable was derived by selecting the first 3 choices made by each respondent and creating a matrix for each respondent based on the first, second and third choice mentioned by them. There were 9 cases (where the respondent selected three languages learnt at the same time. Close to 300 (1.3%) gave two languages equally learnt at the same time.

Recoding ISCOM (Mother) and ISCOF (Father)

The International Standard Classification of Occupation (ISCO) was capture for both the father (F) and mother (M) of the respondent. Because of International requirements these variables were expanded from 2-digits to 4-digits. Two zeros were simply added to the existing 2-digit code. There was a value for "Not applicable" in the Father and Mother's occupation coding in the International File, that needed to be derived and since there was a separate questions that established if the respondent would be able to answer questions about each parent when the respondent was 16 years, they received a not applicable status for the appropriate ISCO code. Thus a code was created for "Not-applicable" using information from the C3A for mother and C7A for father of respondent.

Recoding other Industry and Occupation codes

Values for "don't know" and "refused" for the respondents (Occupation and Industry) were picked up from the original questions as captured in the field, before the file was sent for coding the Occupation and Industry codes. ISCO (International Standard Classification of Occupation), SOC91 (Standard Occupation Classification, 1991) & NOCS (National Occupation Classification for Statistics, 2001) were codes created for Occupation using questions D27 and D28; and NAICS (North American Industrial Classification System), SIC80 (Standard Industrial Classification, 1980) & ISIC (International Standard Industrial Classification) were codes created for Industry, using question D25 and D26.

Recoding of Income questions

There was one extra code "respondent can only report net (after tax) earnings" in questions D41 (Salary and wages before taxes) and D42 (how were you paid per month, gross) in the International file. In order to accommodate this code in the National file, a code "9999999995" was created, if valid values existed in D41A & D42A respectively, i.e. values > 0 and < 9999999995.

Changing variable Length

Some variables were creating with 2-bytes instead of 1-byte. There were many such variables, especially month variables that came from the field, because they were mark-all-questions, and had months 1-12 to accommodate. They were during the Process made 2-digits when de-stringing them, because of the nature of question. Since they only had 1, 2, 6, 7, 8, 9 values, they only needed to be 1-byte long. They were therefore, all made 1-digit questions. D5MA - D5MI; D6AMA - D6AML; D6BMA - D6BML; D10MA – D10ML; D12MA - D12ML; D14MA - D14ML; F5MA - F5ML are some examples.

Also, D42 & D42A in the Labor Force Activities module were changed from 9.2 to 10.2 digits, in order to be consistent with income questions.

All decimals in the National file were removed, except the weight variables.

Creating a unique respondent identification number

A 5-digit IDNO number was generated by the system, such that no record can be identified by geographic location or the person itself. In the International file "sampleid" was removed and only IDNO is used to link International and National file.

6.4.8 Creation of the national Public Use Microdata File (Step 8)

Once the Master file has been finalised, it is possible to construct the Canadian IALSS Public Use Microdata file, 2003. From this second file stems the ALL International Public Use Microdata file.

Thus, the National PUMF is a subset of the Master in terms of the variables released on the file and the level of detail presented.

Statistics Canada employs every method available to protect the confidentiality of each and every respondent to each and every survey it conducts. Of course, the most restrictive of these is the complete absence of a public use microdata file. In the case of the IALSS, it would be difficult to identify a respondent based solely on a prose proficiency score of "283.7839". On the other hand, the more variables one adds, the more unique a respondent appears. They no longer enjoy the anonymity of numbers and re-emerge as an individual. For instance, take as an illustration a male aged 24-35 of Ukrainian decent, University Educated, living in rural Saskatchewan working as a Veterinarian in an agricultural establishment employing over 75 people and an annual salary over $47,000. He is married with a child under 5, takes courses and seminars throughout the year, uses computers daily as part of his work and has a prose proficiency score of "283.7839". Suddenly, we seem to know the "individual" better.

A combination of any large number of uncorrelated variables is bound to produce unique records to appear in the crosstabulation. Every single respondent on the IALSS PUMF has an associated sample weight. Implicit in this weighting is that each respondent is not unique in the population as a whole. In other words, unique records resulting from a combination of variables can only tell us that the respondent was unique within the IALSS sample, but not necessarily unique within the entire population. This is particularly true for respondents with larger weights. The inherent risk of disclosure is greater for the data of respondents with smaller weights. When measures are applied to protect confidentiality of the respondent, additional care was taken for records with smaller weights. These records include respondents from the North and respondents with characteristics targeted by some of the supplementary samples.

It should be pointed out that the analysis to uncover disclosure risks was not limited to an examination of the weights, indeed, several measures were taken to assess risk. Based on the results of these tests and measures, the IALSS adopted various methods designed to protect the confidentiality of its respondents. These include:

1) **Variable suppression**: In certain instances, the sensitivity of individual variables where such that they required complete suppression from the Public Microdata file. The address and telephone number of the respondent falls into this category. Variable suppression was also used to rid the PUMF of administrative variables of questionable analytical value.

2) **Recoding and/or collapsing**: In other instances, the variable, while sensitive in its detail, could be recoded into less visible groupings (income, year of immigration, wages, age, etc…) or, the existing groupings could be collapsed into less visible larger groups. While this was clearly an attempt to include the most information as possible, such data concerns where clearly subjugated to the higher goal of confidentiality. In other words, an attempt was made to keep as many variables as possible, with as much detail as possible, but protection of confidentiality was the overriding concern in the preparation of the file.

3) **Local suppression**: In between the two options, local suppression involves the complete suppression of a variable, but only for selected respondents or identifiable group of respondents. This preserves the details for the bulk of the population with only marginal impacts on the standard errors of estimates used with these variables.

In essence, variables that did not pose a confidentiality risk were written to the IALSS PUMF as is. Those that posed limited risk for only certain sub-populations have had local suppressions imposed (the values have been suppressed for a select group of individual respondents). Those that posed a more general risk due to their detail where either recoded or had their values collapsed into groupings that were large enough to mask any confidentialty issue. Finally, those that posed a clear confidentiality risk that could not be otherwise addressed or those that provided little or no analytical value, have been entirely suppressed from the IALSS public use micro data file.

6.4.9 International PUMF creation (step 9)

Just as the national PUMF can be viewed as a subset of the National Master File, the Canadian component of the ALL International Public Use Microdata file is, in effect, a subset of the Canadian National Public Use Microdata File. Variables included on the National file but not on the International PUMF include aboriginal status, immigration status, mother tongue and second language usage, official language proficiency, Canadian education groupings and other variables that are either idiosyncratic to Canada or that have no comparable counterpart with other countries.

7.0 Guidelines for tabulation and analysis

This section of the documentation outlines the guidelines to be adhered to by users tabulating, analysing, publishing or otherwise releasing any data derived from the survey microdata tapes. With the aid of these guidelines, users of microdata should be able to produce the same figures as those produced by Statistics Canada and, at the same time, will be able to develop currently unpublished figures in a manner consistent with these established guidelines.

7.1 Sample weighting guidelines for tabulation

The IALSS surveys are based upon complex sample designs, with stratification, multiple stages of selection, and unequal probabilities of selection of respondents. Using data from such complex surveys presents problems to analysts because the survey design and the selection probabilities affect the estimation and variance calculation procedures that should be used. In order for survey estimates and analyses to be free from bias, the survey weights must be used.

While many analysis procedures found in statistical packages allow weights to be used, the meaning or definition of the weight in these procedures differ from that which is appropriate in a sample survey framework, with the result that while in many cases the estimates produced by the packages are correct, the variances that are calculated are poor. Programs for calculating standard errors for simple estimates such as totals, proportions and ratios (for qualitative variables) are provided in the following section.

7.2 Definitions of types of estimates: Categorical vs. Quantitative

Before discussing how the IALSS data can be tabulated and analyzed, it is useful to describe the two main types of point estimates of population characteristics, which can be, generated from the microdata file for the IALSS.

Categorical estimates:

Categorical estimates are estimates of the number, or percentage of the surveyed population possessing certain characteristics or falling into some defined category. The number of Albertans at literacy Level 1 on the prose scale or the proportion of Canadians at literacy Level 4 in numeracy are examples of such estimates. An estimate of the number of persons possessing a certain characteristic may also be referred to as an estimate of an aggregate.

Examples of Categorical Questions:

Q: Do you ever watch television or videos in a language other than French or English?
R: **Yes / No**

Q: How would you rate your reading skills in English needed in daily life?
R: **Excellent / Good / Moderate / Poor**

Quantitative estimates:

Quantitative estimates are estimates of totals or of means, medians and other measures of central tendency of quantities based upon some or all of the members of the surveyed population. They also specifically involve estimates of the form $\hat{X}/\hat{Y}$ where $\hat{Y}$ is an estimate of surveyed population quantity total and $\hat{Y}$ is an estimate of the number of persons in the surveyed population contributing to that total quantity.

An example of a quantitative estimate is the average number of employers that working Canadians had in the past 12 months. The numerator is an estimate of the total number of employers that working Canadians had in the past 12 months, and its denominator is the number of Canadians reporting that they worked in the past 12 months.

Examples of Quantitative Questions :

Q: How many different employers have you had in the past 12 months?
R: **|_|_| employer(s)**

Q: How many hours per week did you usually work at this job?
R: **|_|_| hours**

7.2.1 Tabulation of categorical estimates

Estimates of the number of people within a given country with a certain characteristic can be obtained from the microdata file by summing the final weights of all records possessing the characteristic(s) of interest.

Proportions and ratios of the form $\hat{X}/\hat{Y}$ for a country are obtained by:

1) summing the final weights of records having the characteristic of interest for the numerator ($\hat{X}$),
2) summing the final weights of records having the characteristic of interest for the denominator ($\hat{Y}$), then
3) dividing the numerator estimate by the denominator estimate.

7.2.2 Tabulation of quantitative estimates

Estimates of quantities can be obtained from the microdata file by multiplying the value of the variable of interest by the final weight for each record, then summing this quantity over all records of interest. For example, to obtain an estimate for a particular country of the total number of different employers that people working part time have had in the past 12 months, multiply the value reported in the question D4 (number of employers) by the final weight for the record, then sum this value over all records with D5=2 (part time).

To obtain a weighted average of the form $\hat{X}/\hat{Y}$, the numerator ($\hat{X}$) is calculated as for a quantitative estimate and the denominator ($\hat{Y}$) is calculated as for a categorical estimate. For example, to estimate the average number of employers in the past 12 months of people working part time, in a given country:

a) estimate the total number of employers as described above,
b) estimate the number of people in this category by summing the final weights of all records with QD5=2, then
c) divide estimate a) by estimate b).

7.3 Skill level estimates

The IALSS design is an adaptation of a three parameter (PL) Item Response Theory model. The first parameter (A) is the ability of the item to discriminate (sensitivity to proficiency) and the second (B) is its difficulty. A third parameter (C) is the lower asymptote parameter which reflects the possibly non-zero chance of a correct response independent of ability. However, since the IALSS test did not generally use any multiple choice type questions, this (C) parameter was fixed at zero throughout, thus transforming the equation into what can now be called a 2PL model. Once the parameters have been calculated, each item can be assigned a Response Probability value of 80 (RP80) which measures the proficiency level needed for a respondent to answer the task with an 80% probability of success.

As noted previously, a respondent's proficiency in the three scales was summarized through the use of the item parameters and the respondent's ability in accordance with the IRT scaling models. The application differed from the norm in that the IALSS called for administering relatively few items to each respondent in order to track population levels of proficiency more efficiently. Because the data are not intended to estimate individual levels of proficiency, however, more complicated analyses are required.

Plausible values methodology was used to estimate key population features consistently and to approximate others no less accurately than standard IRT procedures would. In essence, this added dimension requires that the estimation of proficiency be based on a series of five plausible values for each of the three literacy domains. These five plausible values—prose1 through prose5 for the prose scale, doc1 through doc5 for the document scale and num1 through num5 for the numeracy scale and health1 through health5 for the health Literacy scale—have been recoded into plausible levels with values from 1 through 5 reflecting the empirically determined progression of information-processing skills and strategies required to perform increasingly complex tasks. Level 1 is equivalent to scores in the range 0 to 226 (inclusive); Level 2 is equal to scores of 226.0001 through 276; Level 3 goes from 276.0001 to 326; Level 4 includes scores ranging from 326.0001 to 376 and, Level 5 is equivalent to scores greater or equal to 376.0001. For the prose scale, the variables are called plev1 through plev5, for the document scale, these are dlev1 through dlev5 and for the numeracy scale, nlev1 through nlev5.

Due to a difference in the framework, the Problem Solving scale was treated slightly differently. First, the same five plausible values ranging from 0-500 where created (prob1 through prob5), but the level definition was slightly different. For instance, this scale only defines 4 levels of proficiency with level 1 being the weakest and level 4 the highest level of proficiency. Thus, while it is necessairy to collapse levels 4 and 5 in order to replicate the published estimates (there are typically too few respondents at level 5 to produce reliable estimates) for the prose, document, numeracy and health scales, this step is not required for the Problem solving domain.

For simple point estimates in either of the five skill domains, it is often sufficient to use the population weight along with one of the corresponding five plausible values (chosen at random).

However, a more precise point estimate can be obtained by taking the average of the five estimates produced from each of the five plausible values, which can be computed as follows:

$\boldsymbol{T_{.} = (\sum_i T_i) / 5}$, where T_i is a vector of five weighted estimates from each of the five plausible values.

Note that taking an average of the five plausible values, will only produce a valid point estimate, not a valid variance estimate. **All five** plausible values as well as the 30 replicate weights must be used in order to correctly compute design-based variance estimates. Design-based variance estimates are discussed further in section 8.1.2. (Using Plausible Values and Replicate Weights in Calculating Sampling Errors).

7.4 Rounding guidelines

In order that estimates for publication or other release derived from the microdata file correspond to those produced by Statistics Canada, users are urged to adhere to the following guidelines regarding the rounding of such estimates:

a) Estimates in the main body of a statistical table are to be rounded to the nearest hundred units using the normal rounding technique. In normal rounding, if the first or only digit to be dropped is 0 to 4, the last digit to be retained is not changed. If the first or only digit to be dropped is 5 to 9, the last digit to be retained is raised by one. For example, in normal rounding to the nearest 100, if the last two digits are between 00 and 49, they are changed to 00 and the preceding digit (the hundreds digit) is left unchanged. If the last digits are between 50 and 99 they are changed to 00 and the preceding digit is incremented by 1.

b) Marginal sub-totals and totals in statistical tables are to be derived from their corresponding unrounded components and then are to be rounded themselves to the nearest 100 units using normal rounding.

c) Averages, proportions, rates and percentages are to be computed from unrounded components (i.e. numerators and/or denominators) and then are to be rounded themselves to one decimal using normal rounding. In normal rounding to a single digit, if the final or only digit to be dropped is 0 to 4, the last digit to be retained is not changed. If the first or only digit to be dropped is 5 to 9, the last digit to be retained is increased by 1.

d) Sums and differences of aggregates (or ratios) are to be derived from their corresponding unrounded components and then are to be rounded themselves to the nearest 100 units (or the nearest one decimal) using normal rounding.

e) In instances where, due to technical or other limitations, a rounding technique other than normal rounding is used resulting in estimates to be published or otherwise released which differ from corresponding estimates published by Statistics Canada, users are urged to note the reason for such differences in the publication or release document(s).

f) Under no circumstances are unrounded estimates to be published or otherwise released by users. Unrounded estimates imply greater precision than actually exists.

8.0 Data quality

The data quality from any survey can be evaluated by looking at two types of survey errors: sampling error and non-sampling error.

The estimates derived from this survey are based on a sample of individuals. Somewhat different figures might have been obtained if a complete census had been taken using the same questionnaire, interviewers, supervisors, processing methods, etc. as those actually used. The difference between the estimates obtained from the sample and the results from a complete count taken under similar conditions is called the **sampling error** of the estimate.

Errors, which are not related to sampling, may occur at almost every phase of a survey operation. Interviewers may misunderstand instructions, respondents may make errors in answering questions, the answers may be incorrectly entered on the questionnaire and errors may be introduced in the processing and tabulation of the data. These are all examples of **non-sampling errors.**

8.1 Sampling errors

Since it is an unavoidable fact that estimates from a sample survey are subject to sampling error, sound statistical practice calls for researchers to provide users with some indication of the magnitude of this sampling error. This section of the documentation outlines the measures of sampling error which Statistics Canada commonly uses and which it urges users producing estimates from this microdata file to use also.

The basis for measuring the potential size of sampling errors is the standard error of the estimates derived from survey results.

However, because of the large variety of estimates that can be produced from a survey, the standard error of an estimate is usually expressed relative to the estimate to which it pertains. This resulting measure, known as the coefficient of variation (C.V.) of an estimate, is obtained by dividing the standard error of the estimate by the estimate itself and is expressed as a percentage of the estimate.

For example, suppose that, based upon the survey results, one estimates that 16.6% of Canadians are at literacy Level 1 with regard to prose, and this estimate is found to have standard error of 0.013. Then the coefficient of variation of the estimate is calculated as:

$$\left(\frac{.013}{.166}\right) x\ 100\% \ =\ 7.8\%$$

8.1.1 CV release guidelines

One criterion that can used to determine whether survey estimates are publishable is the coefficient of variation (CV). The CV is the standard error of an estimate expressed as a percentage of that estimate.

Before releasing and/or publishing any estimate from the IALS, users should first determine the quality level of the estimate. The quality levels are acceptable, marginal and unacceptable. Data quality is affected by both sampling and non-sampling errors. However for release purposes, the quality level of an estimate will be determined only on the basis of sampling error as reflected by the coefficient of variation as shown in table 8.1. Nonetheless users should be sure to read section 8 to be more fully aware of the quality characteristics of these data.

First, the number of respondents who contribute to the calculation of the estimate should be determined. If this number is less than 30, the weighted estimate should be considered to be of unacceptable quality. For weighted estimates based on sample sizes of 30 or more, users should determine the coefficient of variation of the estimate and follow the guidelines below. These quality level guidelines should be applied to weighted rounded estimates. All estimates can be considered releasable. However, those of marginal or unacceptable quality level must be accompanied by a warning to caution subsequent users.

Table 8.1 **Quality level guidelines**

Quality level of estimate	Guidelines
1. Acceptable	Estimates have: a sample size of 30 or more, and low coefficients of variation in the range 0.0% to 16.5%. No warning is required.
2. Marginal	Estimates have: a sample size of 30 or more, and high coefficients of variation in the range 16.6% to 33.3%. Estimates should be flagged with the letter M (or some similar identifier). They should be accompanied by a warning to caution subsequent users about the high levels of error associated with the estimates.
3. Unacceptable	Estimates have: a sample size of less than 30, or very high coefficients of variation in excess of 33.3%. Statistics Canada recommends not to release estimates of unacceptable quality. However, if the user chooses to do so then estimates should be flagged with the letter U (or some similar identifier) and the following warning should accompany the estimates: “The user is advised that . . . (specify the data) . . . do not meet Statistics Canada’s quality standards for this statistical program. Conclusions based on these data will be unreliable, and most likely invalid. These data and any consequent findings should not be published. If the user chooses to publish these data or findings, then this disclaimer must be published with the data.”

8.1.2 Using plausible values and replicate weights in calculating sampling errors

The following section has been liberally copied from the documentation that accompanies the STATTOOL (SAS and SPSS) programs designed by Statistics Canada to help users manipulate the ALL/IALSS data. The programs and tools discussed in this section are included with the IALSS Public Use Microdata file CD-ROM under the directory called “STATTOOL”. While some details of the following section may be more particular to international comparisons of the type facilitated by the ALL Public Use Microdata file, the discussion that ensues will shed some light on the proper usages and practical limits of the IALSS data as well.

Calculating point estimates

In this section, we will see how to use the sampling weights (POPWT) to obtain population estimates such as percentages (totals) and means (Calculation of standard errors will be presented in section 6).

All examples will be based on a fictional population with the following characteristics:

Taking the easy way out for preliminary analysis

There is little doubt that the IALSS dataset is difficult to manipulate. The 5 Plausible values for the 5 domains (if you include health literacy) along with the 30 replicate weights make the procedures for accurate assessment of standard errors a convoluted affair.

In many instances, simplification of the process, particularly at the exploratory stage would greatly cut down on the processing time required to output the skill estimate analysis.

For this reason, it is recommended that preliminary research use only one of the Plausible values, rather than all five. This is much more accurate than averaging the five plausible values, since it allows for the weighted population distribution to accurately reflect the point estimate. The average of the PV would mask the testing error and, as the population under investigation gets smaller, the estimates will increasingly diverge from the true population distribution.

Of course, once the research is ready for publication, the replicate weights and 5 plausible values should be used to produce the final estimates with accurate standard errors. A full description of this procedure can be found beginning in section 8.1.2.3.

Type	Gender	Population distribution	Sample Distribution (unweighted)	Distribution (weighted)
		%	%	%
Rural	Male	40	30	38
	Female	60	70	62
	Total	**20**	**50**	**19**
Urban	Male	51	45	50
	Female	49	55	50
	Total	**80**	**50**	**81**
Total	Male	48.8	37.5	47.7
	Female	51.2	62.5	52.3
	Total	**100.0**	**100.0**	**100.0**

From this table, it seems that the male participation was lower then the female participation in both rural and urban areas. Even though nearly 49% of the population is made of males, we have only 37.5% males in the sample. This trend can be observed in both areas. It seems also that the rural area was over allocated with 50% of the sample coming from that area compared to only 20% in the population.

However, once the sampling weights are used, the percentages are quite comparable. How are they calculated?

Percentages (Totals)

The weighted percentage of the males living in rural areas was calculated as follow:

$$Weighted\ \% = \frac{\sum_{i=1}^{rural,male} POPWT_i}{\sum_{i=1}^{rural} POPWT_i} = 38\%$$

where i identifies individual i. The numerator is an estimate of the total population of males living in rural areas while the denominator is an estimate of the total population living in the rural areas.

The unweighted percentage was calculated as follow:

$$Unweighted\ \% = \frac{\sum_{i=1}^{rural,male} 1_i}{\sum_{i=1}^{rural} 1_i} = \frac{n_{rural,male}}{n_{rural}} = 30\%$$

where $n_{rural,male}$ is the total number of males living in rural areas found in the sample and n_{rural} is the total number of people living in rural areas found in the sample.

In the latter, each sampled individual accounts for one while in the weighted version, each sampled unit was given a weight in order to properly and proportionally represent the subgroups in the sample (note that the weighted percentage is a ratio of estimated weighted totals).

Means

For this fictional example, let's say that we also have the average score based on variable PROSE1 as illustrated by the following table:

Type	Gender	Population distribution	Sample			
			Distribution (unweighted)	Distribution (weighted)	Avg. Prose1 (unweighted)	Avg. Prose1 (weighted)
		%	%	%		
Rural	Male	40	30.0	38.0	260.0	260.1
	Female	60	70.0	62.0	29.00	289.8
	Total	**20**	**50.0**	**19.0**	**281.0**	**278.5**
Urban	Male	51	45.0	50.0	320.0	319.7
	Female	49	55.0	50.0	330.0	330.1
	Total	**80**	**50.0**	**81.0**	**325.5**	**324.9**
Total	Male	48.8	37.5	47.7	296.0	310.7
	Female	51.2	62.5	52.3	307.6	321.0
	Total	**100.0**	**100.0**	**100.0**	**303.3**	**316.1**

Here again we see that the weighted means are quite close to the unweighted means as long as one controls by area type. This is not true for the last 3 lines of the table. Let's try to see why. The weighted mean of the males living in rural areas was calculated as follow:

$$Weighted\ mean = \frac{\sum_{i=1}^{rural,male} POPWT_i * PROSE1_i}{\sum_{i=1}^{rural,male} POPWT_i} = 260.1$$

where i identifies individual i. The numerator is an estimate of the total score for all males living in rural areas while the denominator is an estimate of the total male population living in rural areas.

The unweighted mean was calculated as follow:

$$Unweighted\ mean = \frac{\sum_{i=1}^{rural,male} PROSE1_i}{\sum_{i=1}^{rural,male} 1_i} = \frac{\sum_{i=1}^{rural,male} PROSE1_i}{n_{rural,male}} = 260$$

The unweighted and weighted results will be similar whenever values of PROSE1 don't vary much from one individual to the other and/or values of POPWT behave the same way. This statement doesn't hold for the last three lines of the table. The weighted mean for male is obtained by solving the following equation:

$$Weighted\ mean = \frac{\sum_{i=1}^{male} POPWT_i * PROSE1_i}{\sum_{i=1}^{male} POPWT_i}$$

$$= \frac{(38\% * 19\% * 260.1) + (50\% * 81\% * 319.7)}{47.7\%} = 310.7$$

While the unweighted mean is given by:

$$Unweighted\ mean = \frac{\sum_{i=1}^{male} PROSE1_i}{\sum_{i=1}^{male} 1_i}$$

$$= \frac{(30\% * 50\% * 260) + (45\% * 50\% * 320)}{37.5\%} = 296.0$$

In the latter, each sampled individual accounts for one while in the weighted version, each sampled unit was given a weight in order to properly and proportionally represent the subgroups in the sample. For example, the 30%X50%=15% of males living in rural areas found in the sample was adjusted by the weights to account for 38%X19%=7.22% of the entire sample which is a much better reflect of what is found in the whole population (Note that the true population proportion of males living in rural areas is 40%X20%=8%).

In conclusion, any statistics computed form sample data should always be done using the sampling weights.

Alternative sampling weights

As we saw earlier, the sum of the sampling weights under POPWT within a sample provides an estimate of the size of the population. Although this is a commonly used sampling weight, it sometimes adds to a very large number, and to different numbers from country to country. This is not always desirable. For example, if you want to compute a weighted estimate of the mean achievement in the population across all countries (or sub-populations within a country), using the variable POPWT as your weight variable will lead each country to contribute proportionally to its population size, with the larger countries counting more than small countries. In general, POPWT is not the weight of choice for cross-country analyses. Another consequence of using POPWT is the tendency to inflate results in significance tests when computer softwares are unable to deal correctly with weighted data. We will now see two possible versions of individual sampling weights that address these issues in particular. These versions take advantage on the fact that the same population estimates for means and proportions is obtained whenever you use a weight variable proportional to the population weight (POPWT).

Sum to Constant Sampling Weight (CONSTWT)

It is possible to modify the population weight POPWT such that all countries would contribute the same in a cross-country mean or proportion. This is given by:

$$CONSTWT_{g,i} = POPWT_{g,i} * \left[\frac{100}{\sum_{i=1}^{g} POPWT_i} \right]$$

for each individual in the group of interest g. The transformation of the weights will be different within each country, but in the end the sum of the variable CONSTWT within each country will be 100. The variable CONSTWT, within each country, is proportional to POPWT multiplied by the ratio of 100 divided by the sum of weights over all individuals in the group of interest. These weights can be used when international estimates are sought and you want to have each country contribute the same amount to the international estimate, regardless of the size of the group of interest in the country (see table below).

Group of interest	Country	Population count (rural)	Population estimates	
			Mean PROSE1 (POPWT)	Mean PROSE1 (CONSTWT)
Rural	A	3 700 000	290	290
	B	37 000 000	260	260
	C	7 000 000	300	300
Overall			268	283

Sum to Sample Size Sampling Weight (SMPLWT)

It is possible to modify the population weight POPWT when you want the actual sample size to be used in performing significance tests (within each country). This is given by:

$$SMPLWT_{g,i} = POPWT_{g,i} * \left[\frac{n_g}{\sum_{i=1}^{g} POPWT_i} \right]$$

for each individual in the group of interest g where n_g is the actual sample size in group g. The transformation of the weights will be different within each country, but in the end the sum of the variable CONSTWT within each country will add up to the sample size in group g. The variable SMPLWT, within each country, is proportional to POPWT multiplied by the ratio of the sample size (n_g) divided by the sum of weights over all individuals in the group of interest. Although some statistical computer software packages allow you to use the sample size as the divisor in the computation of standard errors, others will use the sum of the weights, and this results in severely deflated standard errors for the statistics if POPWT is used as the weighting variable. When performing analyses using such software, it is recommended to use a weighting variable such as SMPLWT as the weight variable. Because of the clustering effect in most country samples, it may also be desirable to apply a correction factor such as a design effect to the SMPLWT variable.

Using the plausible values to compute point estimates

To achieve its goal of broad coverage of the literacy purposes and processes, the IALSS /assessment included a range of items arranged into assessment booklets. Each individual participating in the assessment completed one booklet keeping individual response burden to a minimum. ALL used a matrix-sampling design to assign assessment booklets to individuals so that a comprehensive picture of the literacy achievement in each country could be assembled from the components completed by each individual. ALL relied on Item Response Theory (IRT) scaling to combine the individual responses to provide accurate estimates of literacy achievement in the population in each country. The ALL IRT scaling also uses multiple imputation or “plausible values” methodology to obtain proficiency scores in literacy for all individuals, even though each individual responded to only a part of the assessment item pool.

Most cognitive skills testing is concerned with accurately assessing the performance of individual respondents for the purposes of diagnosis, selection or placement. The accuracy of these measurements can be improved by increasing the number of items given to the individual. For the distribution of proficiencies in large population, however, more efficient estimates can be obtained from matrix-sampling design. These designs solicit few responses from each sampled respondent while maintaining a wide range of content representation when responses are aggregated across all respondents. With this approach, however, the advantage of estimating population characteristics is more efficiently offset by the inability to make precise statements about individuals, with the result that aggregations of individual scores can lead to seriously biased estimates of population characteristics.

Plausible values methodology was developed as a way to address this issue by using all available data to estimate directly the characteristics of populations and sub-populations,and then generating multiple imputed scores (called plausible values) from these distributions, which can be used in analyses with standard statistical software. A detailed review of plausible values methodology is given by Mislevy (1991). The main things to retain from this are:

a) Whenever you want to compute statistics involving scores (like PROSE, DOC, NUMERACY, etc) you don't have one score value but five score values assigned to each individuals. Each set of plausible values is equally well-designed to estimate population parameters;

b) These statistics based on scores should always be computed at population or subpopulation levels. They should never be used to do inference at individual level.

Working with plausible values

Example1: Estimated median for the variable PROSE in Country A.
For each individual, we don't have one but 5 scores to deal with as illustrated in the next table

Country A	PROSE1	PROSE2	PROSE3	PROSE4	PROSE5
Individual 1	222	275	300	245	254
Individual 2	289	310	212	250	265
…	…	…	…	…	…
Individual n	285	275	243	321	312
Median	285	281	283	279	289

In order to estimate the overall median for PROSE in country A we first have to estimate the median based on the first set of plausible values found under variable PROSE1. We then repeat this first step using PROSE2 through PROSE5 to get a total of five equally good estimates of the median for that country. Since they are all equally good, the next step to do to obtain a single estimate of the median is to average out these five estimates. We get:

Overall Median = (285 + 281 + 283 + 279 + 289) / 5 = 283.4
Note that you should not average out scores at the individual level. For example, (222 + 275 + 300 + 245 + 254) / 5 = 259.2 is not a good estimate of the variable PROSE for individual 1. This is true for any of the PROSEn variables for a given individual. Score variables should always be interpreted in populations or subpopulations context.

PROSE1 through PROSE5 are the raw ability scores. When these scores are re-grouped into levels, they yield five level variables called for the prose domain PLEV1 through PLEV5 on the PUMF. The variables XPROSE1 through XPROSE5 are a recode of the PLEV1 through PLEV5 variables found on the IALSS PUMF such that levels 4 and 5 are collapsed in order to allow sufficient numbers of respondents in each level for accurate analyse. The same can be duplicated for the Document, Numeracy and Health level variables (DLEV1-DLEV5, NLEV1-

NLEV5, HLEV1-HLEV5). The Problem Solving levels should be left as found in PSLEV1 through PSLEV5.

Example 2: Estimated logistic regression parameter coefficients (Levels 2 versus level 3 of the dependent variables XPROSE1 to XPROSE5).

From the values found in the previous table this would give:

Country A	XPROSE1	XPROSE2	XPROSE3	XPROSE4	XPROSE5
Individual 1	2	3	3	2	2
Individual 2	3	4/5	2	2	2
…	…	…	…	…	…
Individual n	3	2	2	3	3

The first thing we note is that a given individual may be found in different proficiency levels depending on which set of plausible values we are looking at. This does not invalidate the methodology used however. As explained before, in order to get the logistic regression parameter coefficients at the country level, we first have to calculate these parameters based on the first set of plausible values. We then repeat this step 4 times using XPROSE2 through XPROSE5 to get 4 additional sets of estimated parameters as illustrated in the next table:

Country A	XPROSE1	XPROSE2	XPROSE3	XPROSE4	XPROSE5
Intercept	0.124	0.129	0.122	0.125	0.120
Beta 1	1.051	1.059	1.049	1.055	1.060
Beta 2	0.584	0.591	0.545	0.499	0.645
Beta 3	3.222	4.123	3.012	3.542	3.201

Since these sets of estimated parameter coefficients are all equally good, the next step to do to obtain a single set of estimates is to average out these five results. We get:

Overall Intercept = (0.124 + 0.129 + 0.122 + 0.125 + 0.120) / 5 = 0.124
Overall Beta 1 = (1.051 + 1.059 + 1.049 + 1.055 + 1.060) / 5 = 1.055
Overall Beta 2 = (0.548 + 0.591 + 0.545 + 0.499 + 0.645) / 5 = 0.566
Overall Beta 3 = (3.222 + 4.123 + 3.012 + 3.542 + 3.201) / 5 = 3.420

8.1.3 Estimating error variance in IALSS

The IALSS methodology is a tried and tested method of quantifying skills. As such, a large body of documentation already exists regarding the appropriate ways of estimating variance in such studies. A large part of the text found in this section was borrowed from the IEA PIRLS 2001 User's Guide and adapted to fit the IALSS context.

Overview

When analysing data from complex designs such as IALSS, it is important to compute correct error variance estimates for the statistics of interest. In All this error variance can come from two sources: the sampling process (always present) and the imputation process (whenever the statistics of interest involve proficiency scores). This section describes the methods used to estimate these error variance components.

Estimating sampling variance

When data are collected as part of a complex sample survey, analytically there is often no easy way to produce unbiased or design-consistent estimate of variance. A class of techniques called replication methods provides a way to estimate variance for the type of complex sample designs such as those used in ALL.

The basic idea behind replication is to select subsamples repeatedly from the whole sample, calculate the statistic of interest for each subsample, and then use these subsamples or replicate statistics to estimate the variance of the full-sample statistic. Different ways of creating subsamples from the full-sample result in different replication methods. The subsamples are called replicates and the statistics calculated from these replicates are called replicate estimates.

One such method is the jackknife repeated replication (JRR) technique (Wolter, 1985). In ALL, within each country, the full sample was randomly split into 30 subsets of equal or nearly equal size, with each subset resembling the full sample. Replicates are formed by deleting one subset at a time and multiplying the weights for the other subsets by

$$\frac{30}{29}$$

In this manner, 30 replicates are created. This method is also known as the JK1 method. The weights associated with each replicate can be found under variables REPLIC01 through REPLIC30. These weights should only be used to calculate the sampling variance. Point estimates should be calculated as described in the previous section.

Computing sampling variance using the JK1 method

The basic idea here is to calculate the estimate of interest from the full sample using the weight variable POPWT (or SMPLWT or CONSTWT) as well as each replicate (using the variables REPLIC01 through REPLIC30). The variation between the replicate estimates and the full-sample estimate is then used to estimate the sampling variance for the full sample. Suppose that $\hat{\theta}$ is the full-sample estimate of some population parameter θ. The sampling variance estimator $\mathrm{var}(\hat{\theta})$ is given by:

$$\mathrm{var}_{smpl}(\hat{\theta}) = \frac{30\sum_{g=1}^{30}(\hat{\theta}_{(g)} - \hat{\theta})^2}{29}$$ where $\hat{\theta}_{(g)}$ is the estimate of θ based on the observations included in the g-th replicate.

When the statistic of interest involves proficiency scores, it is common practice to base the sampling variance on the first set of plausible values only rather than computing the above expression 5 times and averaging out the 5 estimated sampling variances.

Example 1: Average personal income in country A

The average personal income is given by the following expression:

$$Mean\ personal\ income = \frac{\sum_{i=1}^{Country\ A} POPWT_i * D43_i}{\sum_{i=1}^{Country\ A} POPWT_i} = 27\ 601$$

In order to compute the sampling variance, we have to calculate the following expression 30 times, each one based on the appropriate replicate weight.

$$Mean\ personal\ income_{(1)} = \frac{\sum_{i=1}^{Country\ A} REPLIC01_i * D43_i}{\sum_{i=1}^{Country\ A} REPLIC01_i} = 26\ 983$$

$$Mean\ personal\ income_{(2)} = \frac{\sum_{i=1}^{Country\ A} REPLIC02_i * D43_i}{\sum_{i=1}^{Country\ A} REPLIC02_i} = 26\ 146$$

• • •

$$Mean\ personal\ income_{(30)} = \frac{\sum_{i=1}^{Country\ A} REPLIC30_i * D43_i}{\sum_{i=1}^{Country\ A} REPLIC30_i} = 28\ 965$$

We then simply apply the variance formula given earlier. This gives:

$$\mathrm{var}_{smpl}(\hat{\theta}) = \frac{29\sum_{g=1}^{30}(\hat{\theta}_{(g)} - \hat{\theta})^2}{30} = \frac{29}{30}\left[(26983 - 27601)^2 + (26146 - 27601)^2 + \cdots + (28965 - 27601)^2\right]$$

Finally, the statistic $(\hat{\theta} - \theta) / \mathrm{var}(\hat{\theta})^{1/2}$ is approximately t-distributed with 29 degrees of freedom.

Estimating imputation variance

Whenever the statistics of interest involve proficiency scores, there is a need for estimating the imputation variance. As mentioned in previous section, five scores are generated for the same test for individuals participating in IALSS. These different scores are referred to as plausible values (PVs). In IALSS, each individual was presented with blocks of exercises. The full collection of blocks covers the concepts to be tested, but an individual respondent did not answer questions from all blocks. Using a type of balanced assignment of block of respondents, the full battery of questions was covered when respondents are aggregated. For a group of similar respondents, a Bayesian posterior distribution of scores was estimated. The plausible values for each respondent are realizations from the posterior distribution. These scores are not meaningful for an individual respondent, but when combined can be used to estimate population averages and other population quantities.

Computing imputation variance

The general procedure for estimating the imputation variance using plausible values is as follows:

a) First estimate the statistic of interest θ, each time using a different set of plausible values (M) and the variable POPWT (or SMPLWT or CONSTWT). Let's call these 5 estimates, $\hat{\theta}_1$ to $\hat{\theta}_5$. The statistic of interest can be anything estimable from the sample data, such as mean, the difference between means, percentiles, regression parameter coefficients, etc.

b) Then estimate the overall estimate by averaging out the $\hat{\theta}_m$ where m=1, 2,…5. Let's call this estimate $\hat{\theta}$.

c) The imputation variance is computed as:

$$Var_{imp}(\hat{\theta}) = \left[1 + \frac{1}{5}\right] \times \sum_{m=1}^{5} \frac{\left(\hat{\theta}_m - \hat{\theta}\right)^2}{4}$$

Estimating the overall error variance

Under ideal circumstances and with unlimited computing resources, the overall error variance would be computed as follows:

$$Var(\hat{\theta}) = \sum_{m=1}^{5} \frac{Var_{smpl}(\hat{\theta}_m)}{5} + Var_{imp}(\hat{\theta})$$

Since each of the Var$_{smpl}$ involves calculating the statistics of interest 30 times (each time using a different sampling weight), this shortcut formula can be used instead where sampling variance is estimated from the first set of plausible values only.

$$Var(\hat{\theta}) = Var_{smpl}(\hat{\theta}_1) + Var_{imp}(\hat{\theta})$$

When the statistics of interest do not involve any proficiency scores, the overall error variance formula simply becomes:

$$Var(\hat{\theta}) = Var_{smpl}(\hat{\theta})$$

Degrees of freedom

$(\hat{\theta} - \theta) / Var(\hat{\theta})^{1/2}$ is approximately t-distributed, with degrees of freedom (Jonhson & Rust,1993) given by:

$$\nu = \frac{1}{\frac{f_m^2}{4} + \frac{(1 - f_m)^2}{29}}$$

where f_m is given by:

$$f_m = \frac{Var_{imp}(\hat{\theta})}{Var(\hat{\theta})}$$

In practice, the number of degrees of freedom is set to 29 (this will work well when f is relatively small, say less than 30%).

Making comparisons

We will now see how to compute the correct error variance when comparing survey estimates between countries, to the international estimates, and within countries. In order to simplify the text, the estimated mean achievement for the variable PROSE will be considered only. It should be straightforward to generalise this section to any type of survey estimate.

Between countries

The error variance when comparing the estimated mean achievement for PROSE between country A and B is given by:

$$Var(\hat{\theta}) = Var_{Country\ A}(\hat{\theta}) + Var_{Country\ B}(\hat{\theta})$$

For example, say that the estimated mean achievement for country A is 290 with an error variance of 25 and for country B, the estimated mean achievement is 307 with an error variance of 30.25. The difference between these two countries is then 307-290 = 17. The question is : is this difference of 17 points the result of error due to sampling only part of the population combined with the fact that only part of the items were administered? To find the answer to that question, we first have to compute the following statistic (known as the Wald statistic):

$$(\hat{\theta}) / Var(\hat{\theta})^{1/2} = 17 / (25 + 30.25)^{1/2} = 2.287$$

When this value is compared to the critical 95% value from a t distribution with 29 degrees of freedom (2.04), we conclude that there is enough evidence to state that these two countries don’t have the same estimated mean achievement.
Note that this approach is also valid when comparing the ALL results to the IALS results.

To the international estimates

An important published statistics shows your country mean achievement compared to the international mean. The error variance when doing so is given by:

$$Var(\hat{\theta}) = \frac{(N-1)^2 Var_{smpl\ A}(\hat{\theta}_A) + \sum_{k=1,k\neq A}^{N} Var_{smpl\ k}(\hat{\theta}_k)}{N^2} + Var_{imp}(\hat{\theta}_A - \hat{\theta}_{int})$$

where N is the number of countries used to compute the international mean, $\hat{\theta}_m$ stands for the estimated mean for country m, and $\hat{\theta}_{int}$ stands for the estimated international mean.

For example, let's consider the following tables:

	Mean achievement	Sampling variance
Country A	290	20
Country B	300	22
Country C	286	18
Country D	324	22
International	300	

Mean achievement	PROSE1	PROSE2	PROSE3	PROSE4	PROSE5
Country A	288	292	292	288	290
International	301	300	299	302	298
$\hat{\theta}_A - \hat{\theta}_{int}$	-13	-8	-7	-14	-8

Here we have that the estimated mean achievement for country A is 290 with a sampling variance of 20 (imputation variance of 5) and the estimated international mean achievement based on 4 countries is 300. The difference between country A result and the international mean is then 290-300 = -10. The question is : is this difference of 10 points the result of error due to sampling only part of the population combined with the fact that only part of the items were administered? To find the answer to that question, we first have to compute the following statistic:

$$Var(\hat{\theta}) = \frac{(3)^2 20 + (22+18+22)}{4^2} + (1+\frac{1}{5})\frac{[(13-10)^2 + (8-10)^2 + (7-10)^2 + (14-10)^2 + (8-10)^2)]}{4}$$

$$= 15.125 + 12.6 = 27.725$$

we then compute the Wald statistics:

$$(\hat{\theta}) / Var(\hat{\theta})^{1/2} = 10/(27.725)^{1/2} = 1.899$$

When this value is compared to the critical 95% value from a t distribution with 29 degrees of freedom (2.04), we conclude that there isn't enough evidence to state that country A is different than the estimated international mean achievement.

Within countries

Most of the times when comparing subgroups within countries, there is no direct formula for computing the overall error variance like we had in the previous two sections. The main reason for this is that, the samples for the different subgroups are not typically treated as independent for the purpose of statistical tests. Accordingly, a jackknife procedure applicable to correlated samples for estimating the sampling variance of the difference between subgroups should be applied. This involves computing the difference between subgroups once for each of the 30 replicate samples, and five more times, once for each set of plausible values as described earlier (see combining sampling and imputation variances).

However, linear regression models can be easily used to compute these differences. Here's how to compute the difference between men and women for the variable PROSE:

a) Create dummy variables for the subgroups; let's call MAN the variable that will take value 1 if the respondent is a man and 0 otherwise, and WOMAN the variable that will take value 1 if the respondent is a woman and 0 otherwise. You create as many dummy variables as there are subgroups.

b) Using POPWT, run a linear regression model with PROSE1 as the dependent variable and either MAN or WOMAN as the independent variable (When there are more k dummy variables with k greater than 2, choose k-1 dummy variables as independent variables). The dummy variable left out will become the reference subgroup.

c) Using the replicate weights, repeat step b), 30 times.

d) Using POPWT, repeat step b) 4 times, once for each set of plausible values

e) Combine information from step b), c), and d) to compute the overall point estimate, the error variance, and the Wald statistics.

For example, say that after creating the dummy variables you get from step b the following result:

Mean PROSE1 = 270 + 18*WOMAN (weighted by POPWT)

This simplifies to Mean PROSE1 = 270 when respondents are males and Mean PROSE1 = 288 when respondents are females. This means that the coefficient in front of the variable WOMAN in the regression model is the difference between women and men while the intercept (270) is the mean achievement for the reference level, men in this case.

From step **c** you get:

Mean PROSE1 = 271 + 17*WOMAN (weighted by REPLIC01)
Mean PROSE1 = 269 + 19*WOMAN (weighted by REPLIC02)
Mean PROSE1 = 273 + 14*WOMAN (weighted by REPLIC03)
…
Mean PROSE1 = 268 + 21*WOMAN (weighted by REPLIC30)

The sampling variance of the difference between women and men can now be calculated as:

$$\text{var}_{smpl}(\hat{\theta}) = \frac{29\sum_{g=1}^{30}(\hat{\theta}_{(g)} - \hat{\theta})^2}{30} = \frac{29}{30}\left[(17-18)^2 + (19-18)^2 + (14-18)^2 + \cdots + (21-18)^2\right] = 19.575$$

From step d you get:

Mean PROSE2 = 271 + 20*WOMAN (weighted by POPWT)
Mean PROSE3 = 269 + 19*WOMAN (weighted by POPWT)
Mean PROSE4 = 273 + 17*WOMAN (weighted by POPWT)
Mean PROSE5 = 268 + 16*WOMAN (weighted by POPWT)

The overall point estimate for the difference can now be computed by averaging out the results over PROSE1 to PROSE5. This gives: (18+20+19+17+16) / 5 = 18. The imputation variance is given by:

$$Var_{imp}(\hat{\theta}) = \left[1 + \frac{1}{5}\right] \times \sum_{m=1}^{5} \frac{\left(\hat{\theta}_m - \hat{\theta}\right)^2}{4}$$

$$= 1.2 \times \frac{\left[(18-18)^2 + (20-18)^2 + (19-18)^2 + (17-18)^2 + (16-18)^2\right]}{4} = 3.0$$

The Wald statistics becomes

$$(\hat{\theta}) / Var(\hat{\theta})^{1/2} = 18/(19.575 + 3.0)^{1/2} = 3.788$$

When this value is compared to the critical 95% value from a t distribution with 29 degrees of freedom (2.04), we conclude that there is enough evidence to state that men mean achievement is different than women mean achievement within country A.

8.1.4 Performing analyses with the IALSS data using SPSS

This section presents some basic examples of analyses that can be performed using the sampling weights and scores discussed in previous sections. It also provides details on selected SPSS programs to conduct such analyses, and the results of these analyses. The analyses presented here are simple in nature. The programs compute the percentage of respondents in specified subgroups, the mean achievement for those groups, and the corresponding standard errors (square root of the total error variance) for the percentage and mean statistics.

In our examples, we use macros written in SPSS that can be used to perform any of the analyses that are described in this section. These are general procedures that can be used for many purposes, provided you have some basic knowledge of the SPSS macro language. If you have some programming experience in this statistical package, you will be able to make the necessary modifications to the macros to obtain the desired results.

The SPSS macros

The four available SPSS macros are described as follows:

JACKMEAN.SPS

This SPSS macro can be used to compute weighted percentages of respondents within defined groups, and their mean (average) on a specified continuous variable. This macro also computes the JRR sampling variances for the percentages and mean estimates. The variable can be any continuous variable in the file.

JACKMEANPV.SPS

This macro can be used in SPSS to compute weighted percentages of respondents within defined groups, and their mean achievement scores on an achievement scale using plausible values. This macro makes use of the plausible values in computing the mean achievement scores. This macro also computes the jackknife repeated replication (JRR) sampling variances for the percentages of respondents within specified groups, and the JRR and imputation variances for the mean achievement scores. This macro should only be used when multiple plausible values are used in the analyses.

JACKREG.SPS

This macro can be used in SPSS to compute linear regression coefficients and their corresponding standard errors within defined groups. This macro can be used with any variable in the analysis but it does not make use of plausible values.

JACKREGPV.SPS

This SPSS macro can be used to compute linear regression coefficients and their corresponding standard errors when using plausible values as the dependent variables within defined groups.

Means and percentages when plausible values are not involved

This section presents example SPSS code that can be used to compute the standard errors for means and percentages of variables other than plausible values. This code is provided in the form of an SPSS program called **JACKMEAN.SPS** that computes the percentages of respondents within subgroups defined by a set of classification variables, the standard errors of these percentages, the means for the groups on a variable of choice, and the standard errors of these means. The standard errors computed by this SPSS macro are taking into account the IALSS sample design.

When using this macro, you need to specify a set of classification variables, one analysis variable, the number of replicate weights (if this number is the same for different countries, you can merge the country data sets together, otherwise run the analysis country by country), the replicate weights and the population weight that is to be used for the analysis. You will also need to specify the data file that contains the data to be processed.

You need to know some basic SPSS syntax in order to use the macro effectively. First it needs to be included in the program file where it is going to be used. If you are operating in batch mode, then the macro needs to be called in every batch. If you are using SPSS interactively, the macro needs to be called once at the beginning of the session and it will remain active throughout the session. If the session is terminated or restarted at a later time the macro needs to be called once again. Once the macro is included in a specific session, the word "JACKMEAN" should not be used within that program because doing so will call the macro.

This macro has several parameters. These are:

INFILE The name of the data file that contains the variables necessary for the analysis (If the path location is included as part of the file name, the name of the file has to be enclosed in quotes). Include only the cases that are of interest in the analysis (e.g., respondents with missing variables have to be excluded prior to calling the macro).

CVAR This lists the variables that are to be used to classify the respondents in the data file. This can be a single variable, or a list of variables. It is recommended to always include the variable that identifies the country. At least one variable had to be specified (e.g., CNTRID).

DVAR This is the variable for which means are to be computed. Only one variable can be listed here.

NJKZ This indicates the number of replicate weights that where generated in the data file. When you are working with the data for only one country, you should set the NJKZ argument to as many replicates as are needed in the country (when more than one country data set, make sure all data sets have the same number of replicates).

RPWT The replicate weights in the data files, generally REPLIC01 to REPLIC30. The replicate weights need to be specified in the form "REPLIC01 TO REPLIC30".

WGT The sampling weight to be used in the analysis, generally POPWT.

The simplest way to call the macro is by using the conventional SPSS notation for invoking macros. This involves listing the macro name followed by the corresponding list of arguments for the analysis, each separated by a slash. For example, if the macro is called using the following code:

```
Include "c:\jackmean.sps".

Jackmean
          Infile = temp                    /
          Cvar  = cntrid                   /
          Dvar  = d43                      /
          Njkz  = 30                       /
          Rpwt = replic01 to replic30      /
          Wgt   = popwt.
```

It will compute the mean of personal income (D43) and its standard error, within each country, using the variable POPWT as the sampling weight. The data will be read from the system file TEMP.

The file that contains these results is called FINAL and is saved to the default directory being used by SPSS. The variables that are contained in this file are:

Classification variables

Each of the classification variables is kept in the resulting file. There is one unique occurrence for each specific combination of the classification variable categories.

Weight variable

It contains the estimate of the population size of the groups defined by each specific combination of the classification variable categories. In the above example, this variable is called POPWT.

MNX

Contains the means of the variable DVAR for the groups defined by the corresponding combinations of classification variable categories.

MNX_SE

Contains the standard errors of the MNX values computed using the jackknife method.

PCT

Contains the percentages of people in the groups for the classification variable listed last, within the specific combination of the categories defined by the groups initially. In our example, we would obtain the percentages of respondents by country.

PCT_SE

Contains the standard errors of PCT computed using the jackknife method.

The file resulting from using this macro can then be printed using the SPSS procedure of choice.

An example is given below.

```
get file = "x:\IALSS\IALSSDATA.sav"
          / keep = cntrid gendaa2 d43 popwt replic01 to replic30 .

select if (gendaa2=1 or gendaa2=2) and not(missing(d43)).

save outfile = respondent.


include "c:\IALSS\jackmean.sps".

jackmean infile= respondent                    /
           cvar = cntrid gendaa2               /
           dvar = d43                          /
           njkz = 30                           /
           rpwt = replic01 to replic30         /
           wgt = popwt.

print formats     cntrid gendaa2 n (F6.0) popwt (f10.0)  mnx mnx_se pct pct_se (f8.2).

report format=list automatic  / var  = cntrid gendaa2 n popwt mnx mnx_se pct pct_se.


        Gender of
CNTRID  Respondent    N       POPWT       MNX      MNX_SE     PCT      PCT_SE
______  __________  ______  __________  ________  ________  ________  ________

    22          1      170      109777  70949.93   7977.68     60.39      3.03
    22          2      190       72007  45423.49   6084.25     39.61      3.03
```

Means and percentages when plausible values are involved

This chapter presents example SPSS code that can be used to compute the standard errors for mean plausible values and percentages. This code is provided in the form of an SPSS macro called **JACKMEANPV.SPS** that computes the percentages of respondents within subgroups defined by a set of classification variables, the standard errors of these percentages, the means for the groups on one of the achievement scales using plausible values, and the standard errors of these means. The standard errors computed by this SPSS macro are taking into account the ALL sample design and the imputation variance components.

When using this macro, you need to specify a set of classification variables, the name of the plausible values and how many there are, the number of replicate weights (if this number is the same for different countries, you can merge the country data sets together, otherwise run the analysis country by country), the replicate weights and the population weight that is to be used for the analysis. You will also need to specify the data file that contains the data to be processed.

You need to know some basic SPSS syntax in order to use the macro effectively. First it needs to be included in the program file where it is going to be used. If you are operating in batch mode, then the macro needs to be called in every batch. If you are using SPSS interactively, the macro needs to be called once at the beginning of the session and it will remain active throughout the session. If the session is terminated or restarted at a later time the macro needs to be called once again. Once the macro is included in a specific session, the word "JACKMEANPV" should not be used within that program because doing so will call the macro. This macro has several parameters. These are:

INFILE The name of the data file that contains the variables necessary for the analysis (If the path location is included as part of the file name, the name of the file has to be enclosed in quotes). Include only the cases that are of interest in the analysis (e.g., respondents with missing variables have to be excluded prior to calling the macro).

CVAR This lists the variables that are to be used to classify the respondents in the data file. This can be a single variable, or a list of variables. It is recommended to always include the variable that identifies the country. At least one variable had to be specified (e.g., CNTRID).

PVS These are the plausible values to be used in the analysis. The plausible values need to be specified in the form "Plausible Value 1 TO Plausible Value 5" as in "PROSE1 TO PROSE5". Although in most cases you will want to use all five plausible values, the program will also work when fewer are specified. You should always use at least two plausible values.

NPV This is the number of plausible values that will be used for the analysis. Generally you will want to use all five plausible values for the analysis although under some circumstances fewer can be used (see PVS above).

NJKZ This indicates the number of replicate weights that where generated in the data file. When you are working with the data for only one country, you should set the NJKZ argument to as many replicates as are needed in the country (when more than one country data set, make sure all data sets have the same number of replicates).

RPWT The replicate weights in the data files, generally REPLIC01 to REPLIC30. The replicate weights need to be specified in the form "REPLIC01 TO REPLIC30".

WGT The sampling weight to be used in the analysis, generally POPWT.

The simplest way to call the macro is by using the conventional SPSS notation for invoking macros. This involves listing the macro name followed by the corresponding list of arguments for the analysis, each separated by a slash. For example, if the macro is called using the following code:

```
Include "c:\jackmeanpv.sps".

Jackmeanpv
            Infile = temp                  /
            Cvar  = cntrid gendaa2         /
            PVS  = PROSE1 to PROSE5        /
            NPV = 5                        /
            Njkz  = 30                     /
            Rpwt = replic01 to replic30    /
            Wgt   = popwt.
```

It will compute the mean prose reading achievement and its standard error for males and females within each country, using five plausible values and the variable POPWT as the sampling weight. It will also compute the percentages of males and females within the country, and their corresponding standard errors. The data will be read from the system file TEMP.

The file that contains these results is called FINAL and is saved to the default directory being used by SPSS. The variables that are contained in this file are:

Classification variables

Each of the classification variables is kept in the resulting file. There is one unique occurrence for each specific combination of the classification variable categories.

Weight variable

It contains the estimate of the population size of the groups defined by each specific combination of the classification variable categories. In the above example, this variable is called POPWT.

N

Contains the number of cases in the groups defined by each specific combination of categories for the classification variables.

MNX

Contains the means for the first plausible value for the groups defined by the corresponding combinations of classification variable categories.

MNX_SE

Contains the standard errors of the mean for the first plausible value for the groups computed using the jackknife method. This does not include the imputation error.

PCT

Contains the percentages of people in the groups for the classification variable listed last, within the specific combination of the categories defined by the groups initially. In our example, it is the percentage of males and females within each country.

PCT_SE

Contains the standard errors of PCT computed using the jackknife method.

MNPV

Contains the means of the plausible values for the groups defined by the corresponding combinations of classification variable categories.

MNPV_SE

Contains the standard errors for the mean of the plausible values for the groups computed using the jackknife method. This includes the sampling and the imputation components.

The file resulting from using this macro can then be printed using the SPSS procedure of choice.
An example is given below.

```
get file = "x:\IALSS\IALSSdata.sav"
      / keep = cntrid gendaa2 popwt replic01 to replic30 prose1 to prose5.

select if (gendaa2=1 or gendaa2=2) .

save outfile = respondent.


include "c:\IALSS\jackmeanpv.sps".

jackmeanpv          infile= respondent              /
                    cvar = cntrid gendaa2           /
                    pvs = prose1 to prose5          /
                    npv=5                    /
                    njkz = 30                /
                    rpwt = replic01 to replic30 /
                    wgt = popwt.

print formats cntrid gendaa2 (F2.0) n (F4.0) popwt (f7.0)   mnpv   mnpv_se  mnx  mnx_se
pct pct_se (f6.2).

report format=list automatic margin(1,255)
  / var  = cntrid gendaa2 n popwt mnpv mnpv_se mnx mnx_se pct pct_se.


        Gender of
CNTRID  Respondent  N     POPWT    MNPV   MNPV_SE   MNX    MNX_SE   PCT  PCT_SE
______  __________ ____  _______  ______  _______  ______  ______  ___________

  22         1     1605  1179970  230.41     1.11  230.39     .99   49.53   .06
  22         2     3196  1202504  226.13      .99  226.64     .91   50.47   .06
```

Regression coefficients when plausible values are not involved

This chapter presents example SPSS code that can be used to compute linear regression coefficients and their standard errors. This code is provided in the form of an SPSS macro called **JACKREG.SPS** that computes the multiple correlation between the specified dependent and independent variables, as well as the regression coefficients and their standard errors. The standard errors computed by this SPSS macro are taking into account the ALL sample design.

When using this macro, you need to specify a set of classification variables, the dependent and independent variables, the number of replicate weights (if this number is the same for different countries, you can merge the country data sets together, otherwise run the analysis country by country), the replicate weights and the population weight that is to be used for the analysis. You will also need to specify the data file that contains the data to be processed.

You need to know some basic SPSS syntax in order to use the macro effectively. First it needs to be included in the program file where it is going to be used. If you are operating in batch mode, then the macro needs to be called in every batch. If you are using SPSS interactively, the macro needs to be called once at the beginning of the session and it will remain active throughout the session. If the session is terminated or restarted at a later time the macro needs to be called once again. Once the macro is included in a specific session, the word “JACKREG” should not be used within that program because doing so will call the macro.

This macro has several parameters. These are:

INFILE The name of the data file that contains the variables necessary for the analysis (If the path location is included as part of the file name, the name of the file has to be enclosed in quotes). Include only the cases that are of interest in the analysis (e.g., respondents with missing variables have to be excluded prior to calling the macro).

CVAR This lists the variables that are to be used to classify the respondents in the data file. This can be a single variable, or a list of variables. It is recommended to always include the variable that identifies the country. At least one variable had to be specified (e.g., CNTRID).

XVAR This is a list of independent variables, at least one, that under the linear regression model will be used as predictors of the dependent variable specified in DVAR. These independent variables can be continuous or categorical, or any other type of coded variable.

DVAR This is dependent variable that under the regression model is predicted by the variable or variables specified by the XVAR parameter. Only one variable can be listed.

NJKZ This indicates the number of replicate weights that where generated in the data file. When you are working with the data for only one country, you should set the NJKZ argument to as many replicates as are needed in the country (when more than one country data set, make sure all data sets have the same number of replicates).

RPWT The replicate weights in the data files, generally REPLIC01 to REPLIC30. The replicate weights need to be specified in the form “REPLIC01 TO REPLIC30”.

WGT The sampling weight to be used in the analysis, generally POPWT.

The simplest way to call the macro is by using the conventional SPSS notation for invoking macros. This involves listing the macro name followed by the corresponding list of arguments for the analysis, each separated by a slash. For example, if the macro is called using the following code:

```
Include “c:\jackreg.sps”.

Jackreg
            Infile = temp                 /
            Cvar  = cntrid                /
            Xvar  = regsex                /
            Dvar  = d43                   /
            Njkz  = 30                    /
            Rpwt = replic01 to replic30   /
            Wgt   = popwt.
```

It will compute the regression equation for the variable REGSEX as a predictor of the personal income. The data will be read from the system file TEMP.

The file that contains these results is called REG and is saved to the default directory being used by SPSS. The variables that are contained in this file are:

Classification variables

Each of the classification variables is kept in the resulting file. There is one unique occurrence for each specific combination of the classification variable categories.

Mult_RSQ

The squared multiple correlation coefficient for the model.

SS_Res, SS_Reg, SS_Total

The residual, regression and total sum of squares for the model within each group as defined by the classification variables.

Regression coefficients and standard errors (B## and B##.SE)

These are the regression coefficients for each of the predictor variables in the modeland their corresponding jackknifed standard errors. The coefficient zero (B00) is the intercept for the model. The other coefficients receive a sequential number starting with 01. This sequential number corresponds to the order of the variables in the list of variables specified in the parameter XVAR.

The file resulting from using this macro can then be printed using the SPSS procedure of choice. An example is given below.

```
get file = "x:\IALSS\IALSSdata.sav"
       / keep = cntrid gendaa2 d43 popwt replic01 to replic30.

select if (gendaa2=1 or gendaa2=2) .    compute regsex = gendaa2 - 1.

save outfile = respondent.

include "c:\IALSS\jackreg.sps".

jackreg            infile= respondent              /
                   cvar = cntrid             /
                   xvar = regsex             /
                   dvar = d43                /
                   njkz = 30                 /
                   rpwt = replic01 to replic30 /
                   wgt = popwt.

print formats      cntrid (F2.0) n (F4.0) mult_RSQ (f5.3)
             SS_Total SS_Reg SS_Res (F10.0) B00 B00.SE B01 B01.SE (f6.2).

report format=list automatic margin(1,255)
  / var  = cntrid n Mult_RSQ SS_Total SS_Reg SS_Res B00 B00.SE B01 B01.SE .

CNTRID  N   MULT_RSQ SS_TOTAL  SS_REG   SS_RES   B00    B00.SE    B01    B01.SE
______ ___  ________ ________ ________ ________  ______ _______   ____   ______

  22   360   .023    1.2E+15   2.8E+13  1.18E+15 70950  7977.7  -25526  9560.6
```

In this example, the variable REGSEX is created by subtracting one from the variable GENDAA2. As a result, males receive a code of 0 and females receive a code of 1 on this variable. In this particular model the variable REGSEX is used to predict the values of the variable D43 (personal income). The model becomes

Personal Income = 70950(7978)   for Males,
Personal Income = 70950(7978) – 25526(9561)   for Females.

The numbers in brackets are the standard errors. This means females have on average a personal income that is $25 526 less than the one for males and $9561 is the standard error attached to that estimate.

Regression coefficients when plausible values are involved

This chapter presents example SPSS code that can be used to compute linear regression coefficients using plausible values as the dependent variable and their standard errors. This code is provided in the form of an SPSS macro called **JACKREGPV.SPS** that computes the average multiple correlation between the specified plausible values and independent variables, as well as the regression coefficients and their standard errors. The standard errors computed by this SPSS macro are taking into account the ALL sample design.

When using this macro, you need to specify a set of classification variables, the dependent and independent variables, the number of replicate weights (if this number is the same for different countries, you can merge the country data sets together, otherwise run the analysis country by country), the replicate weights and the population weight that is to be used for the analysis. You will also need to specify the data file that contains the data to be processed.

You need to know some basic SPSS syntax in order to use the macro effectively. First it needs to be included in the program file where it is going to be used. If you are operating in batch mode, then the macro needs to be called in every batch. If you are using SPSS interactively, the macro needs to be called once at the beginning of the session and it will remain active throughout the session. If the session is terminated or restarted at a later time the macro needs to be called once again. Once the macro is included in a specific session, the word "JACKREGPV" should not be used within that program because doing so will call the macro.

This macro has several parameters. These are:

INFILE The name of the data file that contains the variables necessary for the analysis (If the path location is included as part of the file name, the name of the file has to be enclosed in quotes). Include only the cases that are of interest in the analysis (e.g., respondents with missing variables have to be excluded prior to calling the macro).

CVAR This lists the variables that are to be used to classify the respondents in the data file. This can be a single variable, or a list of variables. It is recommended to always include the variable that identifies the country. At least one variable had to be specified (e.g., CNTRID).

XVAR This is a list of independent variables, at least one, that under the linear regression model will be used as predictors of the dependent variable specified by the plausible values. These independent variables can be continuous or categorical, or any other type of coded variable.

ROOTPV This is the prefix used to identify the plausible values for the achievement scale of interest. For example the root of the prose reading plausible values is “PROSE”.

NPV This is the number of plausible values that will be used for the analysis. Generally you will want to use all five plausible values for the analysis although under some circumstances fewer can be used (see PVS above).

NJKZ This indicates the number of replicate weights that where generated in the data file. When you are working with the data for only one country, you should set the NJKZ argument to as many replicates as are needed in the country (when more than one country data set, make sure all data sets have the same number of replicates).

RPWT The replicate weights in the data files, generally REPLIC01 to REPLIC30. The replicate weights need to be specified in the form “REPLIC01 TO REPLIC30”.

WGT The sampling weight to be used in the analysis, generally POPWT.

The simplest way to call the macro is by using the conventional SPSS notation for invoking macros. This involves listing the macro name followed by the corresponding list of arguments for the analysis, each separated by a slash. For example, if the macro is called using the following code:

```
Include “c:\jackregpv.sps”.

Jackregpv
                    Infile    = temp                  /
                    Cvar      = cntrid                /
                    Xvar      = regsex                /
                    Rootpv    = Prose                 /
                    NPV       = 5                     /
                    Njkz      = 30                    /
                    Rpwt      = replic01 to replic30  /
                    Wgt       = popwt.
```

It will compute the regression equation for the variable REGSEX as a predictor of the plausible values in reading prose. The data will be read from the system file TEMP.

The file that contains these results is called REG and is saved to the default directory being used by SPSS. The variables that are contained in this file are:

Classification variables

Each of the classification variables is kept in the resulting file. There is one unique occurrence for each specific combination of the classification variable categories.

Mult_RSQ

The squared multiple correlation coefficient for the model.

SS_Res, SS_Reg, SS_Total

The residual, regression and total sum of squares for the model within each group as defined by the classification variables.

Regression coefficients and standard errors (B## and B##.SE)

These are the regression coefficients for each of the predictor variables in the model and their corresponding jackknifed standard errors (with both sampling and imputation components). The coefficient zero (B00) is the intercept for the model. The other coefficients receive a sequential number starting with 01. This sequential number corresponds to the order of the variables in the list of variables specified in the parameter XVAR.

The file resulting from using this macro can then be printed using the SPSS procedure of choice. An example is given below.

```
get file = "x:\IALSS\IALSSdata.sav"
        / keep = cntrid gendaa2 popwt replic01 to replic30 prose1 to prose5.

select if (gendaa2=1 or gendaa2=2) . compute regsex = gendaa2 - 1.

save outfile = respondent.

include "c:\IALSS\jackregpv.sps".

jackregpv     infile    = respondent                  /
                     cvar    = cntrid                 /
                     xvar    = regsex                 /
                     rootpv = prose                   /
                     npv    = 5                       /
                     njkz   = 30                      /
                     rpwt   = replic01 to replic30    /
                     wgt    = popwt.

print formats       cntrid (F2.0) n (F4.0) mult_RSQ (f5.3)
             SS_Total SS_Reg SS_Res (F12.0) B00 B00.SE B01 B01.SE (f6.2).

report format=list automatic margin(1,255)
        / var  = cntrid n Mult_RSQ SS_Total SS_Reg SS_Res B00 B00.SE B01 B01.SE .


CNTRID  N  MULT_RSQ  SS_TOTAL   SS_REG    SS_RES    B00    B00.SE   B01  B01.SE
______ ___ ________  ________   ______    _______   ___    ______   ___  ______

  22  4801  .002    4395153088 10952856 4384200232  230.41  1.11  -4.27   1.50
```

In this example, the variable REGSEX is created by subtracting one from the variable GENDAA2. As a result, males receive a code of 0 and females receive a code of 1 on this variable. In this particular model the variable REGSEX is used to predict the values of the plausible values reading PROSE. The model becomes

Prose = 230.41 (1.11)   for Males,
Prose = 230.41 (1.11) – 4.27(1.50)   for Females.

The numbers in brackets are the standard errors. This means females have on average a score in prose that is 4.27 less than the one for males and 1.50 is the standard error attached to that estimate.

8.1.5 Performing analyses with the IALSS data using SAS

This section presents some basic examples of analyses that can be performed using the sampling weights and scores discussed in previous sections. It also provides details on a selected SAS program to conduct such analyses, and the results of these analyses. The analyses presented here are simple in nature. The program computes the percentage of respondents in specified subgroups, the mean achievement for those groups, the weighted counts of respondents in specified groups, the estimated percentiles for these groups, the regression and logistic regression coefficients and their corresponding standard errors (square root of the total error variance).

In our examples, we use a macro written in SAS that can be used to perform any of the analyses that are described in this section. These are general procedures that can be used for many purposes, provided you have some basic knowledge of the SAS macro language. If you have some programming experience in this statistical package, you will be able to make the necessary modifications to the macros to obtain the desired results.

The SAS macro

The only SAS available macro is described as follows: **STATTOOL.SAS**

This macro program in SAS can be used to compute several statistics: means, percentiles, frequencies, counts, differences and regressions (standard linear regression, logistic regression and multinomial regression). These statistics are computed within defined groups taking into account the sampling weights. This macro also computes the JRR standard errors with the sampling and imputation components.

Basic analyses: Means, percentages, counts, percentiles, regression coefficients and their standard errors

This chapter presents example SAS code that can be used to compute means, percentages, counts, percentiles, regression coefficients and their standard errors for any type of variable (whether a plausible values or not). This code is provided in the form of an SAS macro called STATTOOL.SAS that computes these statistics for respondents within subgroups defined by a set of any classification variable (based on plausible values or not). The standard errors computed by this SAS macro are taking into account both sampling and imputation components.

When using this macro, you need to specify a set of classification variables, one analysis variable, the number of replicate weights (if this number is the same for different countries, you can merge the country data sets together, otherwise run the analysis country by country), the replicate weights and the population weight that is to be used for the analysis. You will also need to specify the data file that contains the data to be processed.

You need to know some basic SAS syntax in order to use the macro effectively. First it needs to be included in the program file where it is going to be used. If you are operating in batch mode, then the macro needs to be called in every batch. If you are using SAS interactively, the macro needs to be called once at the beginning of the session and it will remain active throughout the session. If the session is terminated or restarted at a later time the macro needs to be called once again. Once the macro is included in a specific session, the string “%STATTOOL” should not be used within that program because doing so will call the macro.

This macro has several parameters. These are:

WGT The sampling weight to be used in the analysis, generally POPWT

RWGT The root of the variables specifying the replicate weights in the data files, generally REPLIC01 to REPLIC30. The replicate weights need to be specified in the form “REPLIC”.

NREP This indicates the number of replicate weights that where generated in the data file. When you are working with the data for only one country, you should set the NREP argument to as many replicates as are needed in the country (when more than one country data set, make sure all data sets have the same number of replicates).

NPV This is the number of plausible values that will be used for the analysis. Generally you will want to use all five plausible values for the analysis although under some circumstances fewer can be used.

STUDY Put the name of the study (ALL).

CNTRYNO This is the country identifier.

INFILE The name of the data file that contains the variables necessary for the analysis (If the path location is included as part of the file name, the name of the file has to be enclosed in quotes). Include only the cases that are of interest in the analysis (e.g., respondents with missing variables have to be excluded prior to calling the macro).

METHOD This is the statistics you would like to produce. METHOD = mean computes the means of the variable of interest. You can also specified “crosstabs” for crosstabulations, “perc” for percentiles, “diff” for differences, “popest” for population counts, “reg” for standard linear regression, “logistic” for logistic regression and “multinomial” for multinomial logistic regression.

DVAR This is the variable for which means are to be computed. Only one variable can be listed here. Put the name of the variable or only the root if the variable of interest is derived form a set of plausible values.

DVARPV Indicates whether or not the DVAR variable is derived from a set of plausible values. This parameter takes on value 1 if the variable of interest is derived from a set of plausible values and 0 otherwise.

BYVAR This lists the variables that are to be used to classify the respondents in the data file. This can be a single variable, or a list of variables. This parameter defines the subgroups for which means of DVAR are requested.

BYVARPV Indicates whether or not the BYVAR variable(s) is(are) derived from a set of plausible values. This parameter takes on value 1 if the variable of interest is derived from a set of plausible values and 0 otherwise.

In addition to these parameters, 3 parameters, CRITER1, CRITER2, and CRITER3 can be used. They contain one SAS programming statement.

The simplest way to call the macro is by using the conventional SAS notation for invoking macros. This involves listing the macro name followed by the corresponding list of arguments for the analysis, each separated by a coma. For example, if the macro is called using the following code:

```
%include “c:\IALSS\stattool.sas”;

%stattool (wgt    =      popwt,
           rwgt   =      replic,
           nrep   =      30,
           npv    =      5,
           study  =      ALL,
           method =      mean,
           infile =      in,
           dvar   =      prose,
           dvarpv =      1,
           byvar  =      gendaa2   age3,
           byvarpv  =    0 0);
```

It will compute the mean achievement in reading prose using all five sets of plausible values and its standard error, within each group defined by the combination of gender and age categories, using the variable POPWT as the sampling weight. The data will be read from the system file TEMP.

The file that contains these results is called FINALB and is saved to the default directory being used by SAS. There is also a HTML file called FINALB and this one is saved on the C drive of your computer under directory TEMP. This file can easily be accessed using EXCEL from MICROSOFT. The variables that are contained in this file are:

Classification variables

Each of the classification variables is kept in the resulting file. There is one unique occurrence for each specific combination of the classification variable categories.

ESTIMATE

Contains the means of the variable DVAR for the groups defined by the corresponding combinations of the classification variable categories.

STANDARD ERROR

Contains the standard errors of the ESTIMATE values computed using the jackknife method, including both sampling and imputation components.

PROB > |T|

Gives the probability that a Student statistics be larger than the absolute value of the observed estimate, within the specific combination of the categories defined by the groups initially.

Two examples are given below.

```
libname in "C:\IALSS\data";
data in;set in.IALSSdata;run;

%stattool(wgt      =     popwt,
          rwgt     =     replic,
 nrep     =     30,                                                              npv
=     5,                                                                  study    =
ALL,                                                                  method   =   mean,
                                                                infile   =       in,
                                                         dvar     =      prose,
          dvarpv   =    1,
          byvar    =     gendaa2   age3,
          byvarpv  =    0 0);
```

Study: ALL : , ,

Estimated Means for prose by domain and gendaa2 age3, Based on 5 sets of Plausible Values and 29 D.F.

Controlling for domain

Obs	Domain	GENDAA2	AGE3	estimate	Standard Error	Prob >|T|
1	ALL	1	1	236.764	1.81853	0
2	ALL	1	2	231.754	1.87374	0
3	ALL	1	3	197.439	4.25769	0
4	ALL	2	1	236.945	2.05748	0
5	ALL	2	2	225.062	1.08269	0
6	ALL	2	3	192.060	2.77765	0

In this example XPROSE1 to XPROSE5 are used as classification variables (they are the Prose PV1 to PV5 each recoded into levels 1 through 5 with levels 4 and 5 collapsed together. We are estimating the mean personal income by level of reading prose.


```
libname in "C:\IALSS\data";

data in;set in.ialssdata;run;

%stattool(wgt     =     popwt,
          rwgt    =     replic,
          nrep    =     30,
          npv     =     5,
          study   =     ALL,
          method  =     mean,
          infile  =     in,
          dvar    =     D43,
          dvarpv  =     0,
          byvar   =     XPROSE,
          byvarpv =     1,
          criter1 =     if d43 < 99999997 );
```

Study: ALL : if d43 < 99999997 , ,
Estimated Means for d43 by domain and xprose, Based on 5 sets of Plausible Values and 29 D.F. Controlling for domain

Obs	Domain	xprose	estimate	Standard Error	Prob > |T|
1	ALL	1	39799.22	6167.01	.000000463
2	ALL	2	66658.33	9378.51	.000000081
3	ALL	3	103063.21	27265.80	.000724651
4	ALL	4	172247.60	60242.53	.007788203

Altering the content of the "method" allows for the production of various other statistics as follows.

perc = produces percentiles,
diff = provides differences,
popest = produces population counts,
reg = generates a standard linear regression,
logistic = for a logistic regression and,
multinomial = for the production of multinomial logistic regression coefficient.

All other parameters remains the same as for the examples illustrated above with standard errors due to sampling and test error calculated appropriately for each measure.

8.2 Non-sampling errors

Over a large number of observations, randomly occurring non-sampling errors will have little effect on estimates derived from the survey. However, errors occurring systematically will contribute to biases in the survey estimates. Considerable time and effort was made to reduce non-sampling errors in the survey. Quality assurance measures were implemented at each step of the data collection and processing cycle to monitor the quality of the data. These measures included the use of highly skilled interviewers, extensive training of interviewers with respect to the survey procedures and questionnaire, observation of interviewers to detect problems of questionnaire design or misunderstanding of instructions, procedures to ensure that data capture errors were minimized and coding and edit quality checks to verify the processing logic.

Despite these efforts, non-sampling error is bound to exist in every survey. The following text outlines the most likely sources of this error and its impact on the IALSS survey.

8.2.1 Sampling frame

The use of the 2001 Census insured that the IALSS frame was as inclusive as possible and that any exclusions could be effectively calculated into the overall survey design.

8.2.2 Non-response

A major source of non-sampling errors in surveys is the effect of non-response on the survey results. The extent of non-response varies from partial non-response (failure to answer just one or some questions) to total non-response.

Total non-response occurred when the interviewer was either unable to contact the respondent, no member of the household was able to provide the information, or the respondent refused to participate in the survey. The national non-response rate for the IALSS was around 34%. As described in Section 5.9.1, non-response weighting adjustments were performed to compensate for total non-response. The weighting adjustments were calculated within weighting classes formed by using frame information in the case of non-respondent households (no screener data), and by using screener information in the case of non-respondent individuals (screener completed but no data for the selected respondent). These adjustments were designed to reduce the non-response bias as much as possible with the data that were available.

Partial non-response to the survey occurred, in most cases, when the respondent did not understand or misinterpreted a question, refused to answer a question, or could not recall the requested information. Generally, the extent of partial non-response was small in the IALSS.

8.2.3 Response error

A number of other potential sources of non-sampling error that are unique to the IALSS deserve comment. Firstly, some of the respondents may have found the test portion of the study intimidating and this may have had a negative affect on their performance. Unlike “usual” surveys, the IALSS test items have “right” and “wrong” answers. Also, for many respondents this would have been their first exposure to a “test” environment in a considerable number of years. Further, although interviewers did not enforce a time limit for answering questions, the reality of having someone watching and waiting may have, in fact, imposed an unintentional time pressure. It is recognized, therefore that even though items were chosen to closely reflect everyday tasks, the test responses might not fully reveal the literacy capabilities of respondents due to the testing environment. Further, although the test nature of the study called for respondents to perform the activities completely independently of others, situations in the real world often enable persons to sort through printed materials with family, friends and associates. It could be therefore, that the skills measured by the survey do not reflect the full range of some respondents' abilities in a more natural setting.

8.2.4 Scoring

Another potential source of non-sampling error for the IALSS relates to the scoring of the test items, particularly those that were scored on a scale (e.g. items that required respondents to write). Special efforts such as centralizing the scoring and sample verification were made to minimize the extent of scoring errors.

9.0 Coefficients of variance tables

As discussed, the IALSS design is complicated and there is no doubt that the preferred method for assessing the quality of estimates drawn from the data is to use the STATTOOL programs or some equivalent estimation program that can use all 5 plausible values and all 30 replicate weights.

A much less precise method would be to calculate Coefficients of Variance for the estimates, or, when appropriate, to use the tables provided at the end of this chapter.

9.1 Release cut-off's for the International Adult Literacy and Skills Survey

The following table provides an indication of the precision of population estimates as it shows the release cut-offs associated with each of the three quality levels presented in Section 8.1.1. These cut-offs are derived from the coefficient of variation (CV) tables discussed in Section 9.9.

For example, the table shows that the quality of a weighted estimate of 10,000 people possessing a given characteristic in the Atlantic Provinces is marginal.

Note that these cut-offs apply to estimates of population totals only. To estimate ratios, users should not use the numerator value (nor the denominator) in order to find the corresponding quality level. Rule 4 in Section 9.3 and Example 4 in Section 9.4 explain the correct procedure to be used for ratios.

Region	**Acceptable CV 0.0% to 16.5%**		**Marginal CV 16.6% to 33.3%**			**Unacceptable CV > 33.3%**	
Atlantic Provinces	26,500	& over	6,600	to <	26,500	under	6,600
Quebec	93,100	& over	23,100	to <	93,100	under	23,100
Ontario	218,500	& over	54,600	to <	218,500	under	54,600
Western Provinces	82,800	& over	20,500	to <	82,800	under	20,500
Northern Territories	1,300	& over	300	to <	1,300	under	300
Canada	**134,900**	**& over**	**33,300**	**to <**	**134,900**	**under**	**33,300**

9.2 Approximate sampling variability tables

In order to supply coefficients of variation (CV) which would be applicable to a wide variety of categorical estimates produced from this microdata file and which could be readily accessed by the user, a set of Approximate Sampling Variability Tables has been produced. These CV tables allow the user to obtain an approximate coefficient of variation based on the size of the estimate calculated from the survey data.

The coefficients of variation are derived using the variance formula for simple random sampling and incorporating a factor which reflects the multi-stage, clustered nature of the sample design. This factor, known as the design effect, was determined by first calculating design effects for a wide range of characteristics and then choosing from among these a conservative value (usually the 75th percentile) to be used in the CV tables which would then apply to the entire set of characteristics.

The table below shows the conservative value of the design effects as well as sample sizes and population counts by province which were used to produce the Approximate Sampling Variability Tables for the International Adult Literacy and Skills Survey.

Region	**Design effect**	**Sample size**	**Population**
Atlantic Provinces	1.82	4,682	1,890,046
Quebec	1.79	4,166	5,994,042
Ontario	3.13	4,946	9,621,290
Western Provinces	2.07	6,657	7,336,591
Northern Territories	1.67	2,587	59,872
Canada	3.42	23,038	24,901,841

All coefficients of variation in the Approximate Sampling Variability Tables are approximate and, therefore, unofficial. Estimates of actual variance for specific variables may be obtained from Statistics Canada on a cost-recovery basis or by using the 30 jackknife replicate weights, as described in Section 8.1. Since the approximate CV tends to be conservative, the use of actual variance estimates may cause the estimate to be switched from one quality level to another. For instance, a **marginal** estimate could become **acceptable** based on the exact CV calculation.

Remember: If the number of observations on which an estimate is based is less than 30, the weighted estimate is most likely unacceptable and Statistics Canada recommends not to release such an estimate, regardless of the value of the coefficient of variation.

Note: The design effects of estimates for Francophones in the Western provinces is particularly high. Approximate CVs for this subpopulation should be multiplied by a factor of 1.7.

9.3 How to use the coefficient of variation tables for categorical estimates

The following rules should enable the user to determine the approximate coefficients of variation from the Approximate Sampling Variability Tables for estimates of the number, proportion or percentage of the surveyed population possessing a certain characteristic and for ratios and differences between such estimates.

Rule 1: Estimates of numbers of persons possessing a characteristic (Aggregates)

The coefficient of variation depends only on the size of the estimate itself. On the Approximate Sampling Variability Table for the appropriate geographic area, locate the estimated number in the left-most column of the table (headed "Numerator of Percentage") and follow the asterisks (if any) across to the first figure encountered. This figure is the approximate coefficient of variation.

Rule 2: Estimates of proportions or percentages of persons possessing a characteristic

The coefficient of variation of an estimated proportion or percentage depends on both the size of the proportion or percentage and the size of the total upon which the proportion or percentage is based. Estimated proportions or percentages are relatively more reliable than the corresponding estimates of the numerator of the proportion or percentage, when the proportion or percentage is based upon a sub-group of the population. For example, the proportion of females who have completed post-secondary education is more reliable than the estimated number of females who have completed post-secondary education. (Note that in the tables the coefficients of variation decline in value reading from left to right).

When the proportion or percentage is based upon the total population of the geographic area covered by the table, the CV of the proportion or percentage is the same as the CV of the numerator of the proportion or percentage. In this case, Rule 1 can be used.

When the proportion or percentage is based upon a subset of the total population (e.g. those in a particular sex or age group), reference should be made to the proportion or percentage (across the top of the table) and to the numerator of the proportion or percentage (down the left side of the table). The intersection of the appropriate row and column gives the coefficient of variation.

Rule 3: Estimates of differences between aggregates or percentages

The standard error of a difference between two estimates is approximately equal to the square root of the sum of squares of each standard error considered separately. That is, the standard error of a difference $\left(\hat{d} = \hat{X}_1 - \hat{X}_2\right)$ is:

$$\sigma_{\hat{d}} \sqrt{\left(\hat{X}_1 \alpha_1\right)^2 + \left(\hat{X}_2 \alpha_2\right)^2}$$

where $\hat{X}_1$ is estimate 1, $\hat{X}_2$ is estimate 2, and α_1 and α_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively. The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d}$. This formula is accurate for the difference between separate and uncorrelated characteristics, but is only approximate otherwise.

Rule 4: Estimates of ratios

In the case where the numerator is a subset of the denominator, the ratio should be converted to a percentage and Rule 2 applied. This would apply, for example, to the case where the denominator is the number of females and the numerator is the number of females who have completed post-secondary education.

In the case where the numerator is not a subset of the denominator, as for example, the ratio of the number of females who have completed post-secondary education as compared to the number of males who have completed post-secondary education, the standard error of the ratio of the estimates is approximately equal to the square root of the sum of squares of each coefficient of variation considered separately multiplied by $\hat{R}$. That is, the standard error of a ratio $\left(\hat{R} = \hat{X}_1 / \hat{X}_2\right)$ is:

$$\sigma_{\hat{R}} = \hat{R}\sqrt{\alpha_1^{\,2} + \alpha_2^{\,2}}$$

where α_1 and α_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively. The coefficient of variation of $\hat{R}$ is given by $\sigma_{\hat{R}} / \hat{R}$. The formula will tend to overstate the error if $\hat{X}_1$ and $\hat{X}_2$ are positively correlated and understate the error if $\hat{X}_1$ and $\hat{X}_2$ are negatively correlated.

Rule 5: Estimates of differences of ratios

In this case, Rules 3 and 4 are combined. The CVs for the two ratios are first determined using Rule 4, and then the CV of their difference is found using Rule 3.

9.4 Examples of using the coefficient of variation tables for categorical estimates

The following examples based on the IALSS are included to assist users in applying the foregoing rules.

Example 1: Estimates of numbers of persons possessing a characteristic (Aggregates)

Suppose that a user estimates that 5,369,686 females have completed post-secondary education. How does the user determine the coefficient of variation of this estimate?

1) Refer to the coefficient of variation table for Canada in Section 9.9.

2) The estimated aggregate 5,369,686 does not appear in the left-hand column (the "Numerator of Percentage" column), so it is necessary to use the figure closest to it, namely 5,000,000.

3) The coefficient of variation for an estimated aggregate is found by referring to the first non-asterisk entry on that row, namely, 2.4%.

4) So the approximate coefficient of variation of the estimate is 2.4%. The finding that there were 5,369,686 (to be rounded according to the rounding guidelines in Section

7.4) females who have completed post-secondary education is publishable with no qualifications.

Example 2: Estimates of proportions or percentages of persons possessing a characteristic

Suppose that the user estimates that 5,369,686 / 12,668,933 = 42.4% of females have completed post-secondary education. How does the user determine the coefficient of variation of this estimate?

1) Refer to the coefficient of variation table for CANADA in Section 9.9.

2) Because the estimate is a percentage which is based on a subset of the total population (i.e., females who reported their education level), it is necessary to use both the percentage (42.4%) and the numerator portion of the percentage (5,369,686) in determining the coefficient of variation.

3) The numerator, 5,369,686, does not appear in the left-hand column (the "Numerator of Percentage" column) so it is necessary to use the figure closest to it, namely 5,000,000. Similarly, the percentage estimate does not appear as any of the column headings, so it is necessary to use the percentage closest to it, 40.0%.

4) The figure at the intersection of the row and column used, namely 2.1% is the coefficient of variation to be used.

5) So the approximate coefficient of variation of the estimate is 2.1%. The finding that 42.4% of females have completed post-secondary education can be published with no qualifications.

Example 3: Estimates of differences between aggregates or percentages

Suppose that a user estimates that 5,369,686 / 12,668,933 = 42.4% of females have completed post-secondary education, while 5,351,237 / 12,223,350 = 43.8% of males have completed post-secondary education. How does the user determine the coefficient of variation of the difference between these two estimates?

1) Using the CANADA coefficient of variation table in the same manner as described in Example 2 gives the CV of the estimate for women as 2.1%, and the CV of the estimate for men as 2.1% as well.

2) Using Rule 3, the standard error of a difference $\left(\hat{d} = \hat{X}_1 - \hat{X}_2\right)$ is:

$$\sigma_{\hat{d}} = \sqrt{\left(\hat{X}_1\alpha_1\right)^2 + \left(\hat{X}_2\alpha_2\right)^2}$$

where $\hat{X}_1$ is estimate 1 (women), $\hat{X}_2$ is estimate 2 (men), and α_1 and α_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively.

That is, the standard error of the difference $\hat{d} = 0.424 - 0.438 = -0.014$ is:

$$\sigma_{\hat{d}} = \sqrt{[(0.424)(0.021)]^2 + [(0.438)(0.021)]^2}$$
$$= \sqrt{(0.0000793) + (0.0000846)}$$
$$= 0.013$$

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d} = 0.013 / 0.014 = 0.93$

4) So the approximate coefficient of variation of the difference between the estimates is 93%. The difference between the estimates is considered unacceptable and Statistics Canada recommends this estimate not be released. However, should the user choose to do so, the estimate should be flagged with the letter U (or some similar identifier) and be accompanied by a warning to caution subsequent users about the high levels of error associated with the estimate.

Example 4: Estimates of ratios

Suppose that the user estimates that 7,299,911 persons in Ontario read or use information from newspapers at least once a week (G3A=1), while 4,561,049 persons in Ontario read or use information from books at least once a week (G3C=1). The user is interested in comparing the estimate of persons who read newspapers versus persons who read books in the form of a ratio. How does the user determine the coefficient of variation of this estimate?

1) First of all, this estimate is a ratio estimate, where the numerator of the estimate ($\hat{X}_1$) is the number of persons in Ontario who read or use information from newspapers at least once a week. The denominator of the estimate ($\hat{X}_2$) is the number of persons in Ontario who read or use information from books at least once a week.

2) Refer to the coefficient of variation table for Ontario in Section 9.9.

3) The numerator of this ratio estimate is 7,299,911. The figure closest to it is 7,000,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 0.9%.

4) The denominator of this ratio estimate is 4,561,049. The figure closest to it is 5,000,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 1.9%

5) So the approximate coefficient of variation of the ratio estimate is given by Rule 4, which is:

$$\alpha_{\hat{R}} = \sqrt{\alpha_1^{\ 2} + \alpha_2^{\ 2}}$$

where α_1 and α_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively.

That is:

$$\alpha_{\hat{R}} = \sqrt{(0.009)^2 + (0.019)^2}$$
$$= \sqrt{0.000081 + 0.000361}$$
$$= 0.021$$

6) The obtained ratio of persons in Ontario who read or use information from newspapers versus persons in Ontario who read or use information from books is 7,299,911 / 4,561,049 which is 1.60 (to be rounded according to the rounding guidelines in Section 7.4). The coefficient of variation of this estimate is 2.1%, which makes the estimate releasable with no qualifications.

Example 5: Estimates of differences of ratios

Suppose that the user estimates that the ratio of males in Ontario who read or use information from newspapers at least once a week to males in Ontario who read or use information from books at least once a week is 2.08, while it is 1.30 for females. The user is interested in comparing the two ratios to see if there is a statistical difference between them. How does the user determine the coefficient of variation of the difference?

1) First calculate the approximate coefficient of variation for the male ratio ($\hat{R}_1$) and the female ratio ($\hat{R}_2$) as in Example 4. The approximate CV for the male ratio is 5.6% and 4.6% for females.

2) Using Rule 3, the standard error of a difference ($\hat{d} = \hat{R}_1 - \hat{R}_2$) is:

$$\sigma_{\hat{d}} = \sqrt{(\hat{R}_1\alpha_1)^2 + (\hat{R}_2\alpha_2)^2}$$

where α_1 and α_2 are the coefficients of variation of $\hat{R}_1$ and $\hat{R}_2$ respectively. That is, the standard error of the difference $\hat{d}$ = 2.08 – 1.30 = 0.78 is:

$$\sigma_{\hat{d}} = \sqrt{[(2.08)(0.056)]^2 + [(1.30)(0.046)]^2}$$
$$= \sqrt{(0.0136) + (0.00358)}$$
$$= 0.13$$

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d}$ = 0.13 / (0.78) = 0.167.

4) So the approximate coefficient of variation of the difference between the estimates is 16.7%. The estimate is considered of marginal quality. It should be flagged with the letter M (or some similar identifier) and be accompanied by a warning to caution subsequent users about the high levels of error associated with the estimates.

9.5 How to use the coefficient of variation tables to obtain confidence limits

Although coefficients of variation are widely used, a more intuitively meaningful measure of sampling error is the confidence interval of an estimate. A confidence interval constitutes a statement on the level of confidence that the true value for the population lies within a specified range of values. For example a 95% confidence interval can be described as follows:

If sampling of the population is repeated indefinitely, each sample leading to a new confidence interval for an estimate, then in 95% of the samples the interval will cover the true population value.

Using the standard error of an estimate, confidence intervals for estimates may be obtained under the assumption that under repeated sampling of the population, the various estimates obtained for a population characteristic are normally distributed about the true population value. Under this assumption, the chances are about 68 out of 100 that the difference between a sample estimate and the true population value would be less than one standard error, about 95 out of 100 that the difference would be less than two standard errors, and about 99 out of 100 that the difference would be less than three standard errors. These different degrees of confidence are referred to as the confidence levels.

Confidence intervals for an estimate, $\hat{X}$, are generally expressed as two numbers, one below the estimate and one above the estimate, as $\left(\hat{X}-k,\ \hat{X}+k\right)$ where k is determined depending upon the level of confidence desired and the sampling error of the estimate.

Confidence intervals for an estimate can be calculated directly from the Approximate Sampling Variability Tables by first determining from the appropriate table the coefficient of variation of the estimate $\hat{X}$, and then using the following formula to convert to a confidence interval ($CI_{\hat{x}}$):

$$CI_{\hat{x}} = \left(\hat{X} - t\hat{X}\alpha_{\hat{x}},\ \hat{X} + t\hat{X}\alpha_{\hat{x}}\right)$$

where $\alpha_{\hat{x}}$ is the determined coefficient of variation of $\hat{X}$, and

$t = 1$ if a 68% confidence interval is desired;
$t = 1.6$ if a 90% confidence interval is desired;
$t = 2$ if a 95% confidence interval is desired;
$t = 2.6$ if a 99% confidence interval is desired.

Note: Release guidelines which apply to the estimate also apply to the confidence interval. For example, if the estimate is not releasable, then the confidence interval is not releasable either.

Example of using the coefficient of variation tables to obtain confidence limits

A 95% confidence interval for the estimated proportion of females who have completed post-secondary education (from Example 2, Section 9.4) would be calculated as follows:

$\hat{X}$ = 42.4% (or expressed as a proportion 0.424)

t = 2

$\alpha_{\hat{x}}$ = 2.1% (0.021 expressed as a proportion) is the coefficient of variation of this estimate as determined from the tables.

$$CI_{\hat{x}} = \{0.424 - (2)(0.424)(0.021), 0.424 + (2)(0.424)(0.021)\}$$

$$CI_{\hat{x}} = \{0.424 - 0.018, 0.424 + 0.018\}$$

$$CI_{\hat{x}} = \{0.406, 0.442\}$$

With 95% confidence it can be said that between 40.6% and 44.2% of females have completed post-secondary education.

9.6 How to use the coefficient of variation tables to do a T-test

Standard errors may also be used to perform hypothesis testing, a procedure for distinguishing between population parameters using sample estimates. The sample estimates can be numbers, averages, percentages, ratios, etc. Tests may be performed at various levels of significance, where a level of significance is the probability of concluding that the characteristics are different when, in fact, they are identical.

Let $\hat{X}_1$ and $\hat{X}_2$ be sample estimates for two characteristics of interest. Let the standard error on the difference $\hat{X}_1 - \hat{X}_2$ be $\sigma_{\hat{d}}$.

If $t = \dfrac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}}$ is between -2 and 2, then no conclusion about the difference between the characteristics is justified at the 5% level of significance. If however, this ratio is smaller than -2 or larger than +2, the observed difference is significant at the 0.05 level. That is to say that the difference between the estimates is significant.

Example of using the coefficient of variation tables to do a T-test

Let us suppose that the user wishes to test, at 5% level of significance, the hypothesis that there is no difference between the proportion of females who have completed post-secondary education and the proportion of males who have completed post-secondary education. From

Example 3, Section 9.4, the standard error of the difference between these two estimates was found to be 0.013. Hence,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0.424 - 0.438}{0.013} = \frac{-0.014}{0.013} = -1.08$$

Since t = -1.08 is between -2 and 2, then no conclusion about the difference between the characteristics is justified at the 5% level of significance.

9.7 Coefficients of variation for quantitative estimates

For quantitative estimates, special tables would have to be produced to determine their sampling error. Since most of the variables for the IALSS are primarily categorical in nature, this has not been done.

As a general rule, however, the coefficient of variation of a quantitative total will be larger than the coefficient of variation of the corresponding category estimate (i.e., the estimate of the number of persons contributing to the quantitative estimate). If the corresponding category estimate is not releasable, the quantitative estimate will not be either. For example, the coefficient of variation of the total number of hours spent in a program of studies would be greater than the coefficient of variation of the corresponding proportion of persons enrolled in a program of studies. Hence, if the coefficient of variation of the proportion is unacceptable (making the proportion not releasable), then the coefficient of variation of the corresponding quantitative estimate will also be unacceptable (making the quantitative estimate not releasable). Coefficients of variation of such estimates can be derived using the 30 jackknife replicate weights, as described in Section 8.1.

9.8 Coefficients of variation for skill level estimates

As explained in Section 8.1.3.5, the five plausible values used for estimating skill levels add another component to the variance, called the imputation variance. A factor was calculated which should be used to inflate the CVs obtained from the CV tables for estimates of skill level. The inflation factor was determined by first calculating the percentage contribution of the imputation variance to the overall variance for a large number of skill level estimates and then choosing the 75th percentile.

Rule 6: Estimates of skill level

The CVs of skill level estimates are approximated by multiplying the value obtained from the CV table by a factor of 1.4. This rule is to be used in conjunction with Rules 1 to 5 in Section x. For estimates involving more than one variable, Rule 6 is used first to approximate the CV of each skill level variable, and then Rules 1 to 5 are applied to combine the CVs and calculate an overall CV. Two examples are provided below.

Example 6.1: Estimates of proportions with skill levels (Rule 2 and Rule 6)

Suppose that a user estimates that 2,470,476 / 12,230,514 = 20.2% of males obtained a score in the Level 1 prose literacy category. How does the user determine the coefficient of variation of this proportion?

Refer to the coefficient of variation table for CANADA in Section 9.9.

Using Rule 2, the CV obtained from the CV table is 3.8%.

Multiplying 3.8% by the inflation factor of 1.4 gives a CV of 5.3%.

So the approximate coefficient of variation of the proportion is 5.3%, which makes the estimate releasable with no qualifications.

Example 6.2: Estimates of ratios with skill levels (Rule 6 and Rule 4)

Suppose that a user estimates that 4,965,226 persons obtained a score in the Level 1 prose literacy category, while 6,402,351 persons completed less than high school. The user is interested in comparing the estimate of these two quantities in the form of a ratio. How does the user determine the coefficient of variation of this ratio?

1) First of all, this estimate is a ratio estimate, where the numerator of the estimate ($\hat{X}_1$) is the number of persons who obtained a score in the Level 1 prose literacy category. The denominator of the estimate ($\hat{X}_2$) is the number of persons who completed less than high school.

2) Refer to the coefficient of variation table for Canada in Section 9.9.

3) The numerator of this ratio estimate is 4,965,226. The CV obtained from the CV tables for the numerator is 2.4%. Since the numerator is a skill level, Rule 6 applies; the CV is multiplied by a factor of 1.4, which yields an approximate CV for the numerator of 3.4%.

4) The denominator of this ratio estimate is 6,402,351. The CV obtained from the CV tables for the denominator is 2.1%. The denominator is not a skill level so Rule 6 does not apply.

5) The approximate coefficient of variation of the ratio estimate is given by Rule 4, which is:

$$\alpha_{\hat{R}} = \sqrt{\alpha_1^{\,2} + \alpha_2^{\,2}}$$

where α_1 and α_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively. That is:

$$\alpha_{\hat{R}} = \sqrt{(0.034)^2 + (0.021)^2} = 0.040$$

6) So the ratio 4,965,226 / 6,402,351 = 0.78 has an approximate coefficient of variation of 4.0%, which makes the estimate releasable with no qualifications.

9.9 Coefficient of variation tables

International Adult Literacy and Skills Survey, 2003

Approximate Sampling Variability Tables - Canada

NUMERATOR OF PERCENTAGE ('000)	ESTIMATED PERCENTAGE													
	0.1%	1.0%	2.0%	5.0%	10.0%	15.0%	20.0%	25.0%	30.0%	35.0%	40.0%	50.0%	70.0%	90.0%
1	192.1	191.2	190.2	187.3	182.3	177.2	171.9	166.4	160.8	154.9	148.9	135.9	105.3	60.8
2	135.8	135.2	134.5	132.5	128.9	125.3	121.5	117.7	113.7	109.6	105.3	96.1	74.4	43.0
3	110.9	110.4	109.8	108.1	105.3	102.3	99.2	96.1	92.8	89.5	85.9	78.5	60.8	35.1
4	96.0	95.6	95.1	93.7	91.2	88.6	85.9	83.2	80.4	77.5	74.4	67.9	52.6	30.4
5	85.9	85.5	85.1	83.8	81.5	79.2	76.9	74.4	71.9	69.3	66.6	60.8	47.1	27.2
6	78.4	78.1	77.7	76.5	74.4	72.3	70.2	67.9	65.6	63.3	60.8	55.5	43.0	24.8
7	72.6	72.3	71.9	70.8	68.9	67.0	65.0	62.9	60.8	58.6	56.3	51.4	39.8	23.0
8	67.9	67.6	67.3	66.2	64.5	62.6	60.8	58.8	56.8	54.8	52.6	48.0	37.2	21.5
9	64.0	63.7	63.4	62.4	60.8	59.1	57.3	55.5	53.6	51.6	49.6	45.3	35.1	20.3
10	60.7	60.5	60.2	59.2	57.7	56.0	54.4	52.6	50.8	49.0	47.1	43.0	33.3	19.2
11	57.9	57.7	57.4	56.5	55.0	53.4	51.8	50.2	48.5	46.7	44.9	41.0	31.7	18.3
12	55.4	55.2	54.9	54.1	52.6	51.1	49.6	48.0	46.4	44.7	43.0	39.2	30.4	17.5
13	53.3	53.0	52.8	52.0	50.6	49.1	47.7	46.2	44.6	43.0	41.3	37.7	29.2	16.9
14	51.3	51.1	50.8	50.1	48.7	47.4	45.9	44.5	43.0	41.4	39.8	36.3	28.1	16.2
15	49.6	49.4	49.1	48.4	47.1	45.7	44.4	43.0	41.5	40.0	38.4	35.1	27.2	15.7
16	48.0	47.8	47.6	46.8	45.6	44.3	43.0	41.6	40.2	38.7	37.2	34.0	26.3	15.2
17	46.6	46.4	46.1	45.4	44.2	43.0	41.7	40.4	39.0	37.6	36.1	33.0	25.5	14.7
18	45.3	45.1	44.8	44.2	43.0	41.8	40.5	39.2	37.9	36.5	35.1	32.0	24.8	14.3
19	44.1	43.9	43.6	43.0	41.8	40.6	39.4	38.2	36.9	35.5	34.2	31.2	24.1	13.9
20	43.0	42.8	42.5	41.9	40.8	39.6	38.4	37.2	36.0	34.6	33.3	30.4	23.5	13.6
21	41.9	41.7	41.5	40.9	39.8	38.7	37.5	36.3	35.1	33.8	32.5	29.7	23.0	13.3
22	41.0	40.8	40.6	39.9	38.9	37.8	36.6	35.5	34.3	33.0	31.7	29.0	22.4	13.0
23	40.1	39.9	39.7	39.1	38.0	36.9	35.8	34.7	33.5	32.3	31.0	28.3	21.9	12.7
24	39.2	39.0	38.8	38.2	37.2	36.2	35.1	34.0	32.8	31.6	30.4	27.7	21.5	12.4
25	********	38.2	38.0	37.5	36.5	35.4	34.4	33.3	32.2	31.0	29.8	27.2	21.1	12.2
30	********	34.9	34.7	34.2	33.3	32.3	31.4	30.4	29.4	28.3	27.2	24.8	19.2	11.1
35	********	32.3	32.2	31.7	30.8	29.9	29.1	28.1	27.2	26.2	25.2	23.0	17.8	10.3
40	********	30.2	30.1	29.6	28.8	28.0	27.2	26.3	25.4	24.5	23.5	21.5	16.6	9.6
45	********	28.5	28.4	27.9	27.2	26.4	25.6	24.8	24.0	23.1	22.2	20.3	15.7	9.1
50	********	27.0	26.9	26.5	25.8	25.1	24.3	23.5	22.7	21.9	21.1	19.2	14.9	8.6
55	********	25.8	25.7	25.3	24.6	23.9	23.2	22.4	21.7	20.9	20.1	18.3	14.2	8.2
60	********	24.7	24.6	24.2	23.5	22.9	22.2	21.5	20.8	20.0	19.2	17.5	13.6	7.8
65	********	23.7	23.6	23.2	22.6	22.0	21.3	20.6	19.9	19.2	18.5	16.9	13.1	7.5
70	********	22.9	22.7	22.4	21.8	21.2	20.5	19.9	19.2	18.5	17.8	16.2	12.6	7.3
75	********	22.1	22.0	21.6	21.1	20.5	19.8	19.2	18.6	17.9	17.2	15.7	12.2	7.0
80	********	21.4	21.3	20.9	20.4	19.8	19.2	18.6	18.0	17.3	16.6	15.2	11.8	6.8
85	********	20.7	20.6	20.3	19.8	19.2	18.6	18.1	17.4	16.8	16.1	14.7	11.4	6.6
90	********	20.2	20.1	19.7	19.2	18.7	18.1	17.5	16.9	16.3	15.7	14.3	11.1	6.4
95	********	19.6	19.5	19.2	18.7	18.2	17.6	17.1	16.5	15.9	15.3	13.9	10.8	6.2
100	********	19.1	19.0	18.7	18.2	17.7	17.2	16.6	16.1	15.5	14.9	13.6	10.5	6.1
125	********	17.1	17.0	16.8	16.3	15.8	15.4	14.9	14.4	13.9	13.3	12.2	9.4	5.4
150	********	15.6	15.5	15.3	14.9	14.5	14.0	13.6	13.1	12.7	12.2	11.1	8.6	5.0
200	********	13.5	13.5	13.2	12.9	12.5	12.2	11.8	11.4	11.0	10.5	9.6	7.4	4.3
250	********	********	12.0	11.8	11.5	11.2	10.9	10.5	10.2	9.8	9.4	8.6	6.7	3.8
300	********	********	11.0	10.8	10.5	10.2	9.9	9.6	9.3	8.9	8.6	7.8	6.1	3.5
350	********	********	10.2	10.0	9.7	9.5	9.2	8.9	8.6	8.3	8.0	7.3	5.6	3.2
400	********	********	9.5	9.4	9.1	8.9	8.6	8.3	8.0	7.7	7.4	6.8	5.3	3.0
450	********	********	9.0	8.8	8.6	8.4	8.1	7.8	7.6	7.3	7.0	6.4	5.0	2.9
500	********	********	********	8.4	8.2	7.9	7.7	7.4	7.2	6.9	6.7	6.1	4.7	2.7
750	********	********	********	6.8	6.7	6.5	6.3	6.1	5.9	5.7	5.4	5.0	3.8	2.2
1,000	********	********	********	5.9	5.8	5.6	5.4	5.3	5.1	4.9	4.7	4.3	3.3	1.9
1,500	********	********	********	********	4.7	4.6	4.4	4.3	4.2	4.0	3.8	3.5	2.7	1.6
2,000	********	********	********	********	4.1	4.0	3.8	3.7	3.6	3.5	3.3	3.0	2.4	1.4
3,000	********	********	********	********	********	3.2	3.1	3.0	2.9	2.8	2.7	2.5	1.9	1.1
4,000	********	********	********	********	********	********	2.7	2.6	2.5	2.4	2.4	2.1	1.7	1.0
5,000	********	********	********	********	********	********	********	2.4	2.3	2.2	2.1	1.9	1.5	0.9
6,000	********	********	********	********	********	********	********	2.1	2.1	2.0	1.9	1.8	1.4	0.8
7,000	********	********	********	********	********	********	********	********	1.9	1.9	1.8	1.6	1.3	0.7
8,000	********	********	********	********	********	********	********	********	********	1.7	1.7	1.5	1.2	0.7
9,000	********	********	********	********	********	********	********	********	********	********	1.6	1.4	1.1	0.6
10,000	********	********	********	********	********	********	********	********	********	********	********	1.4	1.1	0.6
12,500	********	********	********	********	********	********	********	********	********	********	********	********	0.9	0.5
15,000	********	********	********	********	********	********	********	********	********	********	********	********	0.9	0.5
20,000	********	********	********	********	********	********	********	********	********	********	********	********	********	0.4

NOTE: For correct usage of these tables, please refer to the microdata documentation.

International Adult Literacy and Skills Survey, 2003

Approximate Sampling Variability Tables - Atlantic Provinces

NUMERATOR OF PERCENTAGE ('000)	ESTIMATED PERCENTAGE 0.1%	1.0%	2.0%	5.0%	10.0%	15.0%	20.0%	25.0%	30.0%	35.0%	40.0%	50.0%	70.0%	90.0%
1	85.6	85.2	84.7	83.4	81.2	78.9	76.6	74.1	71.6	69.0	66.3	60.5	46.9	27.1
2	********	60.2	59.9	59.0	57.4	55.8	54.1	52.4	50.6	48.8	46.9	42.8	33.2	19.1
3	********	49.2	48.9	48.2	46.9	45.6	44.2	42.8	41.4	39.8	38.3	34.9	27.1	15.6
4	********	42.6	42.4	41.7	40.6	39.5	38.3	37.1	35.8	34.5	33.2	30.3	23.4	13.5
5	********	38.1	37.9	37.3	36.3	35.3	34.2	33.2	32.0	30.9	29.7	27.1	21.0	12.1
6	********	34.8	34.6	34.1	33.2	32.2	31.3	30.3	29.2	28.2	27.1	24.7	19.1	11.1
7	********	32.2	32.0	31.5	30.7	29.8	28.9	28.0	27.1	26.1	25.1	22.9	17.7	10.2
8	********	30.1	30.0	29.5	28.7	27.9	27.1	26.2	25.3	24.4	23.4	21.4	16.6	9.6
9	********	28.4	28.2	27.8	27.1	26.3	25.5	24.7	23.9	23.0	22.1	20.2	15.6	9.0
10	********	26.9	26.8	26.4	25.7	25.0	24.2	23.4	22.6	21.8	21.0	19.1	14.8	8.6
11	********	25.7	25.6	25.2	24.5	23.8	23.1	22.4	21.6	20.8	20.0	18.3	14.1	8.2
12	********	24.6	24.5	24.1	23.4	22.8	22.1	21.4	20.7	19.9	19.1	17.5	13.5	7.8
13	********	23.6	23.5	23.1	22.5	21.9	21.2	20.6	19.9	19.1	18.4	16.8	13.0	7.5
14	********	22.8	22.6	22.3	21.7	21.1	20.5	19.8	19.1	18.4	17.7	16.2	12.5	7.2
15	********	22.0	21.9	21.5	21.0	20.4	19.8	19.1	18.5	17.8	17.1	15.6	12.1	7.0
16	********	21.3	21.2	20.9	20.3	19.7	19.1	18.5	17.9	17.3	16.6	15.1	11.7	6.8
17	********	20.7	20.6	20.2	19.7	19.1	18.6	18.0	17.4	16.7	16.1	14.7	11.4	6.6
18	********	20.1	20.0	19.7	19.1	18.6	18.0	17.5	16.9	16.3	15.6	14.3	11.1	6.4
19	********	********	19.4	19.1	18.6	18.1	17.6	17.0	16.4	15.8	15.2	13.9	10.8	6.2
20	********	********	19.0	18.7	18.2	17.6	17.1	16.6	16.0	15.4	14.8	13.5	10.5	6.1
21	********	********	18.5	18.2	17.7	17.2	16.7	16.2	15.6	15.1	14.5	13.2	10.2	5.9
22	********	********	18.1	17.8	17.3	16.8	16.3	15.8	15.3	14.7	14.1	12.9	10.0	5.8
23	********	********	17.7	17.4	16.9	16.5	16.0	15.5	14.9	14.4	13.8	12.6	9.8	5.6
24	********	********	17.3	17.0	16.6	16.1	15.6	15.1	14.6	14.1	13.5	12.4	9.6	5.5
25	********	********	16.9	16.7	16.2	15.8	15.3	14.8	14.3	13.8	13.3	12.1	9.4	5.4
30	********	********	15.5	15.2	14.8	14.4	14.0	13.5	13.1	12.6	12.1	11.1	8.6	4.9
35	********	********	14.3	14.1	13.7	13.3	12.9	12.5	12.1	11.7	11.2	10.2	7.9	4.6
40	********	********	********	13.2	12.8	12.5	12.1	11.7	11.3	10.9	10.5	9.6	7.4	4.3
45	********	********	********	12.4	12.1	11.8	11.4	11.1	10.7	10.3	9.9	9.0	7.0	4.0
50	********	********	********	11.8	11.5	11.2	10.8	10.5	10.1	9.8	9.4	8.6	6.6	3.8
55	********	********	********	11.3	11.0	10.6	10.3	10.0	9.7	9.3	8.9	8.2	6.3	3.7
60	********	********	********	10.8	10.5	10.2	9.9	9.6	9.2	8.9	8.6	7.8	6.1	3.5
65	********	********	********	10.3	10.1	9.8	9.5	9.2	8.9	8.6	8.2	7.5	5.8	3.4
70	********	********	********	10.0	9.7	9.4	9.2	8.9	8.6	8.2	7.9	7.2	5.6	3.2
75	********	********	********	9.6	9.4	9.1	8.8	8.6	8.3	8.0	7.7	7.0	5.4	3.1
80	********	********	********	9.3	9.1	8.8	8.6	8.3	8.0	7.7	7.4	6.8	5.2	3.0
85	********	********	********	9.1	8.8	8.6	8.3	8.0	7.8	7.5	7.2	6.6	5.1	2.9
90	********	********	********	8.8	8.6	8.3	8.1	7.8	7.5	7.3	7.0	6.4	4.9	2.9
95	********	********	********	********	8.3	8.1	7.9	7.6	7.3	7.1	6.8	6.2	4.8	2.8
100	********	********	********	********	8.1	7.9	7.7	7.4	7.2	6.9	6.6	6.1	4.7	2.7
125	********	********	********	********	7.3	7.1	6.8	6.6	6.4	6.2	5.9	5.4	4.2	2.4
150	********	********	********	********	6.6	6.4	6.3	6.1	5.8	5.6	5.4	4.9	3.8	2.2
200	********	********	********	********	********	5.6	5.4	5.2	5.1	4.9	4.7	4.3	3.3	1.9
250	********	********	********	********	********	5.0	4.8	4.7	4.5	4.4	4.2	3.8	3.0	1.7
300	********	********	********	********	********	********	4.4	4.3	4.1	4.0	3.8	3.5	2.7	1.6
350	********	********	********	********	********	********	4.1	4.0	3.8	3.7	3.5	3.2	2.5	1.4
400	********	********	********	********	********	********	********	3.7	3.6	3.5	3.3	3.0	2.3	1.4
450	********	********	********	********	********	********	********	3.5	3.4	3.3	3.1	2.9	2.2	1.3
500	********	********	********	********	********	********	********	********	3.2	3.1	3.0	2.7	2.1	1.2
750	********	********	********	********	********	********	********	********	********	********	2.4	2.2	1.7	1.0
1,000	********	********	********	********	********	********	********	********	********	********	********	********	1.5	0.9
1,500	********	********	********	********	********	********	********	********	********	********	********	********	********	0.7

NOTE: For correct usage of these tables, please refer to the microdata documentation.

International Adult Literacy and Skills Survey, 2003

Approximate Sampling Variability Tables - Quebec

NUMERATOR OF PERCENTAGE ('000)	ESTIMATED PERCENTAGE 0.1%	1.0%	2.0%	5.0%	10.0%	15.0%	20.0%	25.0%	30.0%	35.0%	40.0%	50.0%	70.0%	90.0%
1	160.3	159.6	158.8	156.4	152.2	147.9	143.5	138.9	134.2	129.3	124.3	113.4	87.9	50.7
2	113.4	112.9	112.3	110.6	107.6	104.6	101.5	98.2	94.9	91.5	87.9	80.2	62.1	35.9
3	92.6	92.2	91.7	90.3	87.9	85.4	82.8	80.2	77.5	74.7	71.7	65.5	50.7	29.3
4	80.2	79.8	79.4	78.2	76.1	74.0	71.7	69.5	67.1	64.7	62.1	56.7	43.9	25.4
5	71.7	71.4	71.0	69.9	68.1	66.1	64.2	62.1	60.0	57.8	55.6	50.7	39.3	22.7
6	********	65.2	64.8	63.8	62.1	60.4	58.6	56.7	54.8	52.8	50.7	46.3	35.9	20.7
7	********	60.3	60.0	59.1	57.5	55.9	54.2	52.5	50.7	48.9	47.0	42.9	33.2	19.2
8	********	56.4	56.1	55.3	53.8	52.3	50.7	49.1	47.5	45.7	43.9	40.1	31.1	17.9
9	********	53.2	52.9	52.1	50.7	49.3	47.8	46.3	44.7	43.1	41.4	37.8	29.3	16.9
10	********	50.5	50.2	49.4	48.1	46.8	45.4	43.9	42.4	40.9	39.3	35.9	27.8	16.0
11	********	48.1	47.9	47.1	45.9	44.6	43.3	41.9	40.5	39.0	37.5	34.2	26.5	15.3
12	********	46.1	45.8	45.1	43.9	42.7	41.4	40.1	38.7	37.3	35.9	32.7	25.4	14.6
13	********	44.3	44.0	43.4	42.2	41.0	39.8	38.5	37.2	35.9	34.5	31.5	24.4	14.1
14	********	42.7	42.4	41.8	40.7	39.5	38.3	37.1	35.9	34.6	33.2	30.3	23.5	13.6
15	********	41.2	41.0	40.4	39.3	38.2	37.0	35.9	34.7	33.4	32.1	29.3	22.7	13.1
16	********	39.9	39.7	39.1	38.0	37.0	35.9	34.7	33.6	32.3	31.1	28.4	22.0	12.7
17	********	38.7	38.5	37.9	36.9	35.9	34.8	33.7	32.6	31.4	30.1	27.5	21.3	12.3
18	********	37.6	37.4	36.9	35.9	34.9	33.8	32.7	31.6	30.5	29.3	26.7	20.7	12.0
19	********	36.6	36.4	35.9	34.9	33.9	32.9	31.9	30.8	29.7	28.5	26.0	20.2	11.6
20	********	35.7	35.5	35.0	34.0	33.1	32.1	31.1	30.0	28.9	27.8	25.4	19.6	11.3
21	********	34.8	34.7	34.1	33.2	32.3	31.3	30.3	29.3	28.2	27.1	24.8	19.2	11.1
22	********	34.0	33.9	33.3	32.4	31.5	30.6	29.6	28.6	27.6	26.5	24.2	18.7	10.8
23	********	33.3	33.1	32.6	31.7	30.8	29.9	29.0	28.0	27.0	25.9	23.7	18.3	10.6
24	********	32.6	32.4	31.9	31.1	30.2	29.3	28.4	27.4	26.4	25.4	23.2	17.9	10.4
25	********	31.9	31.8	31.3	30.4	29.6	28.7	27.8	26.8	25.9	24.9	22.7	17.6	10.1
30	********	29.1	29.0	28.5	27.8	27.0	26.2	25.4	24.5	23.6	22.7	20.7	16.0	9.3
35	********	27.0	26.8	26.4	25.7	25.0	24.3	23.5	22.7	21.9	21.0	19.2	14.9	8.6
40	********	25.2	25.1	24.7	24.1	23.4	22.7	22.0	21.2	20.5	19.6	17.9	13.9	8.0
45	********	23.8	23.7	23.3	22.7	22.0	21.4	20.7	20.0	19.3	18.5	16.9	13.1	7.6
50	********	22.6	22.5	22.1	21.5	20.9	20.3	19.6	19.0	18.3	17.6	16.0	12.4	7.2
55	********	21.5	21.4	21.1	20.5	19.9	19.3	18.7	18.1	17.4	16.8	15.3	11.8	6.8
60	********	********	20.5	20.2	19.6	19.1	18.5	17.9	17.3	16.7	16.0	14.6	11.3	6.5
65	********	********	19.7	19.4	18.9	18.3	17.8	17.2	16.6	16.0	15.4	14.1	10.9	6.3
70	********	********	19.0	18.7	18.2	17.7	17.2	16.6	16.0	15.5	14.9	13.6	10.5	6.1
75	********	********	18.3	18.1	17.6	17.1	16.6	16.0	15.5	14.9	14.3	13.1	10.1	5.9
80	********	********	17.8	17.5	17.0	16.5	16.0	15.5	15.0	14.5	13.9	12.7	9.8	5.7
85	********	********	17.2	17.0	16.5	16.0	15.6	15.1	14.6	14.0	13.5	12.3	9.5	5.5
90	********	********	16.7	16.5	16.0	15.6	15.1	14.6	14.1	13.6	13.1	12.0	9.3	5.3
95	********	********	16.3	16.0	15.6	15.2	14.7	14.3	13.8	13.3	12.7	11.6	9.0	5.2
100	********	********	15.9	15.6	15.2	14.8	14.3	13.9	13.4	12.9	12.4	11.3	8.8	5.1
125	********	********	********	14.0	13.6	13.2	12.8	12.4	12.0	11.6	11.1	10.1	7.9	4.5
150	********	********	********	12.8	12.4	12.1	11.7	11.3	11.0	10.6	10.1	9.3	7.2	4.1
200	********	********	********	11.1	10.8	10.5	10.1	9.8	9.5	9.1	8.8	8.0	6.2	3.6
250	********	********	********	9.9	9.6	9.4	9.1	8.8	8.5	8.2	7.9	7.2	5.6	3.2
300	********	********	********	********	8.8	8.5	8.3	8.0	7.7	7.5	7.2	6.5	5.1	2.9
350	********	********	********	********	8.1	7.9	7.7	7.4	7.2	6.9	6.6	6.1	4.7	2.7
400	********	********	********	********	7.6	7.4	7.2	6.9	6.7	6.5	6.2	5.7	4.4	2.5
450	********	********	********	********	7.2	7.0	6.8	6.5	6.3	6.1	5.9	5.3	4.1	2.4
500	********	********	********	********	6.8	6.6	6.4	6.2	6.0	5.8	5.6	5.1	3.9	2.3
750	********	********	********	********	********	5.4	5.2	5.1	4.9	4.7	4.5	4.1	3.2	1.9
1,000	********	********	********	********	********	********	4.5	4.4	4.2	4.1	3.9	3.6	2.8	1.6
1,500	********	********	********	********	********	********	********	********	3.5	3.3	3.2	2.9	2.3	1.3
2,000	********	********	********	********	********	********	********	********	********	2.9	2.8	2.5	2.0	1.1
3,000	********	********	********	********	********	********	********	********	********	********	********	********	1.6	0.9
4,000	********	********	********	********	********	********	********	********	********	********	********	********	1.4	0.8
5,000	********	********	********	********	********	********	********	********	********	********	********	********	********	0.7

NOTE: For correct usage of these tables, please refer to the microdata documentation.

International Adult Literacy and Skills Survey, 2003

Approximate Sampling Variability Tables - Ontario

NUMERATOR OF PERCENTAGE ('000)	ESTIMATED PERCENTAGE 0.1%	1.0%	2.0%	5.0%	10.0%	15.0%	20.0%	25.0%	30.0%	35.0%	40.0%	50.0%	70.0%	90.0%
1	246.6	245.5	244.2	240.4	234.0	227.4	220.6	213.6	206.4	198.9	191.1	174.4	135.1	78.0
2	174.3	173.6	172.7	170.0	165.5	160.8	156.0	151.1	145.9	140.6	135.1	123.3	95.5	55.2
3	142.4	141.7	141.0	138.8	135.1	131.3	127.4	123.3	119.2	114.8	110.3	100.7	78.0	45.0
4	123.3	122.7	122.1	120.2	117.0	113.7	110.3	106.8	103.2	99.4	95.5	87.2	67.6	39.0
5	110.3	109.8	109.2	107.5	104.7	101.7	98.7	95.5	92.3	88.9	85.5	78.0	60.4	34.9
6	100.7	100.2	99.7	98.2	95.5	92.9	90.1	87.2	84.3	81.2	78.0	71.2	55.2	31.8
7	93.2	92.8	92.3	90.9	88.5	86.0	83.4	80.7	78.0	75.2	72.2	65.9	51.1	29.5
8	87.2	86.8	86.3	85.0	82.7	80.4	78.0	75.5	73.0	70.3	67.6	61.7	47.8	27.6
9	82.2	81.8	81.4	80.1	78.0	75.8	73.5	71.2	68.8	66.3	63.7	58.1	45.0	26.0
10	********	77.6	77.2	76.0	74.0	71.9	69.8	67.6	65.3	62.9	60.4	55.2	42.7	24.7
11	********	74.0	73.6	72.5	70.6	68.6	66.5	64.4	62.2	60.0	57.6	52.6	40.7	23.5
12	********	70.9	70.5	69.4	67.6	65.7	63.7	61.7	59.6	57.4	55.2	50.4	39.0	22.5
13	********	68.1	67.7	66.7	64.9	63.1	61.2	59.3	57.2	55.2	53.0	48.4	37.5	21.6
14	********	65.6	65.3	64.3	62.5	60.8	59.0	57.1	55.2	53.2	51.1	46.6	36.1	20.8
15	********	63.4	63.1	62.1	60.4	58.7	57.0	55.2	53.3	51.4	49.3	45.0	34.9	20.1
16	********	61.4	61.1	60.1	58.5	56.9	55.2	53.4	51.6	49.7	47.8	43.6	33.8	19.5
17	********	59.5	59.2	58.3	56.8	55.2	53.5	51.8	50.1	48.2	46.3	42.3	32.8	18.9
18	********	57.9	57.6	56.7	55.2	53.6	52.0	50.4	48.6	46.9	45.0	41.1	31.8	18.4
19	********	56.3	56.0	55.2	53.7	52.2	50.6	49.0	47.4	45.6	43.8	40.0	31.0	17.9
20	********	54.9	54.6	53.8	52.3	50.9	49.3	47.8	46.2	44.5	42.7	39.0	30.2	17.4
21	********	53.6	53.3	52.5	51.1	49.6	48.1	46.6	45.0	43.4	41.7	38.1	29.5	17.0
22	********	52.3	52.1	51.3	49.9	48.5	47.0	45.5	44.0	42.4	40.7	37.2	28.8	16.6
23	********	51.2	50.9	50.1	48.8	47.4	46.0	44.5	43.0	41.5	39.8	36.4	28.2	16.3
24	********	50.1	49.8	49.1	47.8	46.4	45.0	43.6	42.1	40.6	39.0	35.6	27.6	15.9
25	********	49.1	48.8	48.1	46.8	45.5	44.1	42.7	41.3	39.8	38.2	34.9	27.0	15.6
30	********	44.8	44.6	43.9	42.7	41.5	40.3	39.0	37.7	36.3	34.9	31.8	24.7	14.2
35	********	41.5	41.3	40.6	39.6	38.4	37.3	36.1	34.9	33.6	32.3	29.5	22.8	13.2
40	********	38.8	38.6	38.0	37.0	36.0	34.9	33.8	32.6	31.4	30.2	27.6	21.4	12.3
45	********	36.6	36.4	35.8	34.9	33.9	32.9	31.8	30.8	29.6	28.5	26.0	20.1	11.6
50	********	34.7	34.5	34.0	33.1	32.2	31.2	30.2	29.2	28.1	27.0	24.7	19.1	11.0
55	********	33.1	32.9	32.4	31.6	30.7	29.8	28.8	27.8	26.8	25.8	23.5	18.2	10.5
60	********	31.7	31.5	31.0	30.2	29.4	28.5	27.6	26.6	25.7	24.7	22.5	17.4	10.1
65	********	30.4	30.3	29.8	29.0	28.2	27.4	26.5	25.6	24.7	23.7	21.6	16.8	9.7
70	********	29.3	29.2	28.7	28.0	27.2	26.4	25.5	24.7	23.8	22.8	20.8	16.1	9.3
75	********	28.3	28.2	27.8	27.0	26.3	25.5	24.7	23.8	23.0	22.1	20.1	15.6	9.0
80	********	27.4	27.3	26.9	26.2	25.4	24.7	23.9	23.1	22.2	21.4	19.5	15.1	8.7
85	********	26.6	26.5	26.1	25.4	24.7	23.9	23.2	22.4	21.6	20.7	18.9	14.7	8.5
90	********	25.9	25.7	25.3	24.7	24.0	23.3	22.5	21.8	21.0	20.1	18.4	14.2	8.2
95	********	25.2	25.1	24.7	24.0	23.3	22.6	21.9	21.2	20.4	19.6	17.9	13.9	8.0
100	********	********	24.4	24.0	23.4	22.7	22.1	21.4	20.6	19.9	19.1	17.4	13.5	7.8
125	********	********	21.8	21.5	20.9	20.3	19.7	19.1	18.5	17.8	17.1	15.6	12.1	7.0
150	********	********	19.9	19.6	19.1	18.6	18.0	17.4	16.9	16.2	15.6	14.2	11.0	6.4
200	********	********	********	17.0	16.5	16.1	15.6	15.1	14.6	14.1	13.5	12.3	9.6	5.5
250	********	********	********	15.2	14.8	14.4	14.0	13.5	13.1	12.6	12.1	11.0	8.5	4.9
300	********	********	********	13.9	13.5	13.1	12.7	12.3	11.9	11.5	11.0	10.1	7.8	4.5
350	********	********	********	12.9	12.5	12.2	11.8	11.4	11.0	10.6	10.2	9.3	7.2	4.2
400	********	********	********	12.0	11.7	11.4	11.0	10.7	10.3	9.9	9.6	8.7	6.8	3.9
450	********	********	********	11.3	11.0	10.7	10.4	10.1	9.7	9.4	9.0	8.2	6.4	3.7
500	********	********	********	********	10.5	10.2	9.9	9.6	9.2	8.9	8.5	7.8	6.0	3.5
750	********	********	********	********	8.5	8.3	8.1	7.8	7.5	7.3	7.0	6.4	4.9	2.8
1,000	********	********	********	********	********	7.2	7.0	6.8	6.5	6.3	6.0	5.5	4.3	2.5
1,500	********	********	********	********	********	********	5.7	5.5	5.3	5.1	4.9	4.5	3.5	2.0
2,000	********	********	********	********	********	********	********	4.8	4.6	4.4	4.3	3.9	3.0	1.7
3,000	********	********	********	********	********	********	********	********	********	3.6	3.5	3.2	2.5	1.4
4,000	********	********	********	********	********	********	********	********	********	********	********	2.8	2.1	1.2
5,000	********	********	********	********	********	********	********	********	********	********	********	********	1.9	1.1
6,000	********	********	********	********	********	********	********	********	********	********	********	********	1.7	1.0
7,000	********	********	********	********	********	********	********	********	********	********	********	********	********	0.9
8,000	********	********	********	********	********	********	********	********	********	********	********	********	********	0.9

NOTE: For correct usage of these tables, please refer to the microdata documentation.

International Adult Literacy and Skills Survey, 2003

Approximate Sampling Variability Tables - Western Provinces

NUMERATOR OF PERCENTAGE ('000)	ESTIMATED PERCENTAGE 0.1%	1.0%	2.0%	5.0%	10.0%	15.0%	20.0%	25.0%	30.0%	35.0%	40.0%	50.0%	70.0%	90.0%
1	150.9	150.2	149.5	147.1	143.2	139.2	135.0	130.7	126.3	121.7	116.9	106.8	82.7	47.7
2	106.7	106.2	105.7	104.1	101.3	98.4	95.5	92.5	89.3	86.1	82.7	75.5	58.5	33.8
3	87.1	86.7	86.3	85.0	82.7	80.4	78.0	75.5	72.9	70.3	67.5	61.6	47.7	27.6
4	75.4	75.1	74.7	73.6	71.6	69.6	67.5	65.4	63.2	60.9	58.5	53.4	41.3	23.9
5	67.5	67.2	66.8	65.8	64.1	62.2	60.4	58.5	56.5	54.4	52.3	47.7	37.0	21.4
6	61.6	61.3	61.0	60.1	58.5	56.8	55.1	53.4	51.6	49.7	47.7	43.6	33.8	19.5
7	57.0	56.8	56.5	55.6	54.1	52.6	51.0	49.4	47.7	46.0	44.2	40.3	31.3	18.0
8	********	53.1	52.8	52.0	50.6	49.2	47.7	46.2	44.7	43.0	41.3	37.7	29.2	16.9
9	********	50.1	49.8	49.0	47.7	46.4	45.0	43.6	42.1	40.6	39.0	35.6	27.6	15.9
10	********	47.5	47.3	46.5	45.3	44.0	42.7	41.3	39.9	38.5	37.0	33.8	26.1	15.1
11	********	45.3	45.1	44.4	43.2	42.0	40.7	39.4	38.1	36.7	35.3	32.2	24.9	14.4
12	********	43.4	43.1	42.5	41.3	40.2	39.0	37.7	36.5	35.1	33.8	30.8	23.9	13.8
13	********	41.7	41.5	40.8	39.7	38.6	37.5	36.3	35.0	33.8	32.4	29.6	22.9	13.2
14	********	40.1	39.9	39.3	38.3	37.2	36.1	34.9	33.8	32.5	31.3	28.5	22.1	12.8
15	********	38.8	38.6	38.0	37.0	35.9	34.9	33.8	32.6	31.4	30.2	27.6	21.4	12.3
16	********	37.6	37.4	36.8	35.8	34.8	33.8	32.7	31.6	30.4	29.2	26.7	20.7	11.9
17	********	36.4	36.2	35.7	34.7	33.8	32.8	31.7	30.6	29.5	28.4	25.9	20.1	11.6
18	********	35.4	35.2	34.7	33.8	32.8	31.8	30.8	29.8	28.7	27.6	25.2	19.5	11.3
19	********	34.5	34.3	33.8	32.9	31.9	31.0	30.0	29.0	27.9	26.8	24.5	19.0	11.0
20	********	33.6	33.4	32.9	32.0	31.1	30.2	29.2	28.2	27.2	26.1	23.9	18.5	10.7
21	********	32.8	32.6	32.1	31.3	30.4	29.5	28.5	27.6	26.6	25.5	23.3	18.0	10.4
22	********	32.0	31.9	31.4	30.5	29.7	28.8	27.9	26.9	26.0	24.9	22.8	17.6	10.2
23	********	31.3	31.2	30.7	29.9	29.0	28.2	27.3	26.3	25.4	24.4	22.3	17.2	10.0
24	********	30.7	30.5	30.0	29.2	28.4	27.6	26.7	25.8	24.8	23.9	21.8	16.9	9.7
25	********	30.0	29.9	29.4	28.6	27.8	27.0	26.1	25.3	24.3	23.4	21.4	16.5	9.5
30	********	27.4	27.3	26.9	26.1	25.4	24.7	23.9	23.1	22.2	21.4	19.5	15.1	8.7
35	********	25.4	25.3	24.9	24.2	23.5	22.8	22.1	21.4	20.6	19.8	18.0	14.0	8.1
40	********	23.8	23.6	23.3	22.6	22.0	21.4	20.7	20.0	19.2	18.5	16.9	13.1	7.5
45	********	22.4	22.3	21.9	21.4	20.7	20.1	19.5	18.8	18.1	17.4	15.9	12.3	7.1
50	********	21.2	21.1	20.8	20.3	19.7	19.1	18.5	17.9	17.2	16.5	15.1	11.7	6.8
55	********	20.3	20.2	19.8	19.3	18.8	18.2	17.6	17.0	16.4	15.8	14.4	11.2	6.4
60	********	19.4	19.3	19.0	18.5	18.0	17.4	16.9	16.3	15.7	15.1	13.8	10.7	6.2
65	********	18.6	18.5	18.3	17.8	17.3	16.7	16.2	15.7	15.1	14.5	13.2	10.3	5.9
70	********	18.0	17.9	17.6	17.1	16.6	16.1	15.6	15.1	14.5	14.0	12.8	9.9	5.7
75	********	********	17.3	17.0	16.5	16.1	15.6	15.1	14.6	14.1	13.5	12.3	9.5	5.5
80	********	********	16.7	16.5	16.0	15.6	15.1	14.6	14.1	13.6	13.1	11.9	9.2	5.3
85	********	********	16.2	16.0	15.5	15.1	14.6	14.2	13.7	13.2	12.7	11.6	9.0	5.2
90	********	********	15.8	15.5	15.1	14.7	14.2	13.8	13.3	12.8	12.3	11.3	8.7	5.0
95	********	********	15.3	15.1	14.7	14.3	13.9	13.4	13.0	12.5	12.0	11.0	8.5	4.9
100	********	********	14.9	14.7	14.3	13.9	13.5	13.1	12.6	12.2	11.7	10.7	8.3	4.8
125	********	********	13.4	13.2	12.8	12.4	12.1	11.7	11.3	10.9	10.5	9.5	7.4	4.3
150	********	********	********	12.0	11.7	11.4	11.0	10.7	10.3	9.9	9.5	8.7	6.8	3.9
200	********	********	********	10.4	10.1	9.8	9.5	9.2	8.9	8.6	8.3	7.5	5.8	3.4
250	********	********	********	9.3	9.1	8.8	8.5	8.3	8.0	7.7	7.4	6.8	5.2	3.0
300	********	********	********	8.5	8.3	8.0	7.8	7.5	7.3	7.0	6.8	6.2	4.8	2.8
350	********	********	********	7.9	7.7	7.4	7.2	7.0	6.8	6.5	6.3	5.7	4.4	2.6
400	********	********	********	********	7.2	7.0	6.8	6.5	6.3	6.1	5.8	5.3	4.1	2.4
450	********	********	********	********	6.8	6.6	6.4	6.2	6.0	5.7	5.5	5.0	3.9	2.3
500	********	********	********	********	6.4	6.2	6.0	5.8	5.6	5.4	5.2	4.8	3.7	2.1
750	********	********	********	********	********	5.1	4.9	4.8	4.6	4.4	4.3	3.9	3.0	1.7
1,000	********	********	********	********	********	4.4	4.3	4.1	4.0	3.8	3.7	3.4	2.6	1.5
1,500	********	********	********	********	********	********	********	3.4	3.3	3.1	3.0	2.8	2.1	1.2
2,000	********	********	********	********	********	********	********	********	2.8	2.7	2.6	2.4	1.8	1.1
3,000	********	********	********	********	********	********	********	********	********	********	********	1.9	1.5	0.9
4,000	********	********	********	********	********	********	********	********	********	********	********	********	1.3	0.8
5,000	********	********	********	********	********	********	********	********	********	********	********	********	1.2	0.7
6,000	********	********	********	********	********	********	********	********	********	********	********	********	********	0.6

NOTES:
1. For correct usage of these tables, please refer to the microdata documentation.
2. The design effects of estimates for Francophones in the Western provinces is particularly high. Approximate CVs for this subpopulation should be multiplied by a factor of 1.7.

International Adult Literacy and Skills Survey, 2003

Approximate Sampling Variability Tables - Northern Territories

NUMERATOR OF PERCENTAGE ('000)	ESTIMATED PERCENTAGE													
	0.1%	1.0%	2.0%	5.0%	10.0%	15.0%	20.0%	25.0%	30.0%	35.0%	40.0%	50.0%	70.0%	90.0%
1	*****	*****	19.0	18.7	18.2	17.7	17.2	16.7	16.1	15.5	14.9	13.6	10.5	6.1
2	*****	*****	*****	13.3	12.9	12.5	12.2	11.8	11.4	11.0	10.5	9.6	7.4	4.3
3	*****	*****	*****	*****	10.5	10.2	9.9	9.6	9.3	9.0	8.6	7.9	6.1	3.5
4	*****	*****	*****	*****	9.1	8.9	8.6	8.3	8.0	7.8	7.4	6.8	5.3	3.0
5	*****	*****	*****	*****	8.2	7.9	7.7	7.4	7.2	6.9	6.7	6.1	4.7	2.7
6	*****	*****	*****	*****	*****	7.2	7.0	6.8	6.6	6.3	6.1	5.6	4.3	2.5
7	*****	*****	*****	*****	*****	6.7	6.5	6.3	6.1	5.9	5.6	5.1	4.0	2.3
8	*****	*****	*****	*****	*****	6.3	6.1	5.9	5.7	5.5	5.3	4.8	3.7	2.1
9	*****	*****	*****	*****	*****	*****	5.7	5.6	5.4	5.2	5.0	4.5	3.5	2.0
10	*****	*****	*****	*****	*****	*****	5.4	5.3	5.1	4.9	4.7	4.3	3.3	1.9
11	*****	*****	*****	*****	*****	*****	5.2	5.0	4.9	4.7	4.5	4.1	3.2	1.8
12	*****	*****	*****	*****	*****	*****	*****	4.8	4.6	4.5	4.3	3.9	3.0	1.8
13	*****	*****	*****	*****	*****	*****	*****	4.6	4.5	4.3	4.1	3.8	2.9	1.7
14	*****	*****	*****	*****	*****	*****	*****	4.5	4.3	4.1	4.0	3.6	2.8	1.6
15	*****	*****	*****	*****	*****	*****	*****	*****	4.2	4.0	3.8	3.5	2.7	1.6
16	*****	*****	*****	*****	*****	*****	*****	*****	4.0	3.9	3.7	3.4	2.6	1.5
17	*****	*****	*****	*****	*****	*****	*****	*****	3.9	3.8	3.6	3.3	2.6	1.5
18	*****	*****	*****	*****	*****	*****	*****	*****	*****	3.7	3.5	3.2	2.5	1.4
19	*****	*****	*****	*****	*****	*****	*****	*****	*****	3.6	3.4	3.1	2.4	1.4
20	*****	*****	*****	*****	*****	*****	*****	*****	*****	3.5	3.3	3.0	2.4	1.4
21	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	3.3	3.0	2.3	1.3
22	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	3.2	2.9	2.2	1.3
23	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	3.1	2.8	2.2	1.3
24	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	2.8	2.1	1.2
25	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	2.7	2.1	1.2
30	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1.9	1.1
35	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1.8	1.0
40	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1.7	1.0
45	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	0.9
50	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	0.9

NOTE: For correct usage of these tables, please refer to the microdata documentation

10.0 Record layouts and univariate counts

Please refer to the accompanying document ‘ IALSS_codebook.pdf ‘ for the record layout and univariate counts for the data file.

11.0 Contacts

CTCES Client Services

Client Services, Centre for Education Statistics
Statistics Canada,
150 Tunney's Pasture Driveway
Ottawa ON K1A 0T6
Tel: 1-800-307-3382
Fax : 613-951-4441
E-mail: educationstats@statcan.ca

HRSDC

Client Services
Learning Policy Directorate
Strategic Policy 140 Promenade du Portage
Phase IV, 3rd Floor
Gatineau, Quebec
K1A 0J9
Tel.: (819) 994-4566
Fax: (819) 997-5433
E-mail: NC-SP-PS-LPDR-DGPAR-GD@hrsdc-rhdcc.gc.ca

NLS

Diana Kaan
Senior Policy Officer
National Literacy Secretariat, HRSD
Government of Canada
360 Albert Street, Suite 1510, Ottawa, ON, K1A 0J9
Tel: 613-946-5722
Fax : 613-946-5882
E-mail: diana.kaan@hrsdc-rhdcc.gc.ca
http://www.hrsdc.gc.ca/en/gateways/nav/top_nav/program/nls.shtml