Microdata User Guide

Survey of Canadian Attitudes toward Learning

2006

Statistics Canada Statistique Canada

Canada

Table of Contents

1.0 Introduction 5

2.0 Background 7

3.0 Objectives 9

4.0 Concepts and Definitions 11

5.0 Survey Methodology 13
- 5.1 Sample Selected from the Labour Force Survey **13**
  - 5.1.1 Population Coverage **13**
  - 5.1.2 Sample Design **13**
  - 5.1.3 Sample Size **15**
  - 5.1.4 Sample Rotation **15**
  - 5.1.5 Modifications to the Labour Force Survey Design for the Survey of Canadian Attitudes toward Learning **15**
- 5.2 Sample Selected from the Canadian Community Health Survey **16**
  - 5.2.1 Target Population **16**
  - 5.2.2 Health Regions **16**
  - 5.2.3 Sample Design **17**
  - 5.2.4 Modifications to the Canadian Community Health Survey Design for the Survey of Canadian Attitudes toward Learning **18**
- 5.3 Sample Size by Province/Territory for the Survey of Canadian Attitudes toward Learning **19**

6.0 Data Collection 21
- 6.1 Supervision and Quality Control **21**
- 6.2 Non-response to the Survey of Canadian Attitudes toward Learning **21**

7.0 Data Processing 23
- 7.1 Data Capture **23**
- 7.2 Editing **23**
- 7.3 Coding of Open-ended Questions **23**
- 7.4 Imputation **23**
- 7.5 Creation of Derived Variables **24**
- 7.6 Weighting **24**

8.0 Data Quality 25
- 8.1 Response Rates **25**
- 8.2 Survey Errors **26**
  - 8.2.1 The Frame **26**
  - 8.2.2 Data Collection **26**
  - 8.2.3 Data Processing **27**
  - 8.2.4 Non-response **27**
  - 8.2.5 Measurement of Sampling Error **27**

9.0 Guidelines for Tabulation, Analysis and Release **29**
9.1 Rounding Guidelines **29**
9.2 Sample Weighting Guidelines for Tabulation **29**
9.3 Definitions of Types of Estimates: Categorical and Quantitative **30**
9.3.1 Categorical Estimates **30**
9.3.2 Quantitative Estimates **30**
9.3.3 Tabulation of Categorical Estimates **31**
9.3.4 Tabulation of Quantitative Estimates **31**
9.4 Guidelines for Statistical Analysis **31**
9.5 Coefficient of Variation Release Guidelines **32**
9.6 Release Cut-offs for the Survey of Canadian Attitudes toward Learning **34**

10.0 Approximate Sampling Variability Tables **35**
10.1 How to Use the Coefficient of Variation Tables for Categorical Estimates **36**
10.1.1 Examples of Using the Coefficient of Variation Tables for Categorical Estimates **37**
10.2 How to Use the Coefficient of Variation Tables to Obtain Confidence Limits **41**
10.2.1 Example of Using the Coefficient of Variation Tables to Obtain Confidence Limits **42**
10.3 How to Use the Coefficient of Variation Tables to Do a T-test **43**
10.3.1 Example of Using the Coefficient of Variation Tables to Do a T-test **43**
10.4 Coefficients of Variation for Quantitative Estimates **44**
10.5 Coefficient of Variation Tables **44**

11.0 Weighting **45**
11.1 Weighting Procedures for the Labour Force Survey **45**
11.2 Weighting Procedures for the Canadian Community Health Survey **46**
11.3 Weighting Procedures for the Survey of Canadian Attitudes toward Learning **48**

12.0 Questionnaire **51**

13.0 Record Layout with Univariate Frequencies **53**

1.0 Introduction

The Survey of Canadian Attitudes toward Learning was conducted by Statistics Canada in April and May, 2006 with the cooperation and support of the Canadian Council on Learning. This manual has been produced to facilitate the manipulation of the microdata file of the survey results.

Any question about the data set or its use should be directed to:

Statistics Canada

Client Services
Special Surveys Division
Telephone: 613-951-3321 or call toll-free 1 800 461-9050
Fax: 613-951-4527
E-mail: ssd@statcan.ca

Canadian Council on Learning

Tracey Lavin
#1805-701 West Georgia
Vancouver, British Columbia V7Y 1C6
Telephone: (604) 662-3101 ext. 314
E-mail: tlavin@ccl-cca.ca

2.0 Background

The Survey of Canadian Attitudes toward Learning (SCAL) was sponsored by the Canadian Council on Learning (CCL). The CCL is a national, independent, non-profit corporation with a mandate to promote and support research to improve all aspects of learning across the country and across all walks of life.

Statistics Canada and the CCL will conduct an annual survey on attitudes toward learning in order to assess Canadians' needs, opinions and knowledge concerning learning and education. The SCAL is Canada's first comprehensive source of data on the attitudes of Canadians in regard to education and learning.

The results of this survey will be of considerable interest to those who help to plan the policies and services of current and future governmental and private sector programs and initiatives in the area of learning and education. The survey will help identify knowledge gaps that exist and lead to a deeper understanding of attitudes toward learning among Canada's population.

3.0 Objectives

The goal of the Survey of Canadian Attitudes toward Learning (SCAL) is to find out the views, attitudes and expectations of Canadians on issues pertaining to learning.

The survey focuses on four domains that represent learning themes of central importance to the Canadian Council on Learning (CCL):

1. Early Childhood Learning,
2. Structured Learning,
3. Adult Education and Work and Learning, and
4. Health and Learning.

These themes are complemented by a fifth domain, Social Capital, which will contribute to work on the CCL's Composite Learning Index. The Composite Learning Index is a combination of indicators which measures learning conditions favourable to the economic and social well-being of Canadians.

4.0 Concepts and Definitions

This chapter outlines concepts and definitions of interest to the users. Users are referred to Chapter 12.0 of this document for a copy of the actual survey questionnaire(s) used.

The questionnaire was designed along the themes related to learning and separated into the following modules:

Survey Introduction
All selected respondents were informed about the purpose of the survey, the users and uses of the data, the voluntary nature and confidentiality provisions of the survey as well as permission to share the collected information with the Canadian Council on Learning.

Labour Force (LF)
All selected respondents were asked their age, gender, marital status, highest level of education and labour market activity.

Screening questions (SQ) and Child Roster (CHD)
All selected respondents were asked if they have children, their age, first name, gender, and the child's attendance at a day-care facility, school, or college/university at any time during the last five years. For selected respondents who had more than one child, a table format was presented to accommodate each child's name, gender, age, and whether or not the child lived in the household. The table allowed for the collection of information for a maximum of 10 children.

Adult Learning (AL)
All respondents were asked about participation in work-related education and training. For participants, questions were asked to determine what motivated them to engage in learning. For non-participants, questions addressed some of the perceived and situational barriers to access.

Structured Learning (SL)
Questions on structured learning focused on goals, quality, priorities, and indicators of the formal education system, that is, elementary, secondary and postsecondary education.

Questions related specifically to elementary and secondary school where asked to parents of children older than 9 years of age who had attended a school in the last five years with the exception of parents who also had a child 0 to 8 years. This exclusion was made to allow more focused questions for respondents of young children on early childhood learning.

Goals: What do Canadians believe are the most important goals of formal education?

Quality: How do Canadians rate the quality of formal education with regard to achieving the goals of public education in Canada?

Priorities: What are Canadians' priorities with respect to creating and maintaining a high quality educational system in Canada?

Indicators: What indicators do Canadians use to rate the quality of formal education (e.g., student achievement, school equipments, teachers' qualifications, equity, etc.)?

Goals, quality, priorities, and indicators can change as a function of domain or level of education; therefore, some questions can determine whether any of these factors are level- or domain-specific. This can contribute to a deeper understanding of what Canadians think of the current quality of public education and can identify areas of public education that are in need of improvement.

Early Childhood Learning (EC)
Selected respondents who had children aged 0 to 8 years were asked questions about their awareness and goals pertaining to different aspects of Early Childhood Learning (ECL). Survey questions also distinguished between beliefs about ECL in different contexts.

Awareness: Extensive work has been done with regard to the importance of early childhood learning, but are Canadians aware of the importance of ECL? The survey aimed to assess the level of awareness with regard to learning in the early years.

Goals: What are Canadians' beliefs about the goals for learning in the early years?

Context: Parental and non-parental care (i.e., childcare centres) provide the two main contexts in which young children have the opportunity to learn. The survey examined Canadian beliefs about the quality and accessibility of these contexts as they apply to learning in the early years. In addition, some questions assessed the importance of cooperation between these two main contexts.

Health and Learning (HL)
All respondents were asked questions on how they learned about health, in general; where they received their information about particular health issues; and to a lesser extent, what motivates people to learn about health. The Health and Learning questions sought to identify the ways in which Canadians would like to have more and/or better access to health-related information. This section includes four health challenge scenarios. Each respondent was randomly assigned to one of the health challenges.

Social Capital (SC)
All respondents were asked about the resources available within social groups. Social capital refers to the availability of resources within social groups, such as families, communities, companies, and social clubs. Social capital is the value people derive from their interactions with other people. There is strong evidence that people who have access to social capital are well situated to take full advantage of their learning opportunities. Thus, in monitoring those factors relevant to education and learning that will be included in the Composite Learning Index, it is important to consider the state of social capital among Canadians. The Social Capital questions provide baseline information on social capital in Canada.

Demographic Questions (DM)
In this module, all selected respondents were asked for socio-demographic information about themselves and their families. This includes country of origin, immigration status, date of entry into Canada, ethnic origin, language spoken at home, fluency in the two official languages, number of persons in the household, as well as questions on income and sources of income.

5.0 Survey Methodology

The Survey of Canadian Attitudes toward Learning (SCAL) was administered in April and May, 2006 to a sub-sample of persons selected from the dwellings in the Labour Force Survey (LFS) sample for the 10 provinces, and the Canadian Community Health Survey (CCHS) in the territories. The SCAL sample design is closely tied to these surveys. The LFS design is briefly described in Section 5.1[1], while the design for the CCHS is described in Section 5.2[2]. These sections describe how the SCAL departed from the basic LFS and CCHS design. The initial sample sizes by province and territory for the SCAL are presented in Section 5.3.

5.1 Sample Selected from the Labour Force Survey

5.1.1 Population Coverage

The LFS is a monthly household survey of a sample of individuals who are representative of the civilian, non-institutionalized population 15 years of age or older in Canada's 10 provinces. Specifically excluded from the survey's coverage are residents of the Yukon, Northwest Territories and Nunavut, persons living on Indian Reserves, full-time members of the Canadian Armed Forces and inmates of institutions. These groups together represent an exclusion of approximately 2% of the population aged 15 or over.

5.1.2 Sample Design

The LFS has undergone an extensive redesign, culminating in the introduction of the new design at the end of 1994. The LFS sample is based upon a stratified, multi-stage design employing probability sampling at all stages of the design. The design principles are the same for each province.

Primary Stratification
Provinces are divided into economic regions (ER) and employment insurance economic regions (EIER). ERs are geographic areas of more or less homogeneous economic structure formed on the basis of federal-provincial agreements. They are relatively stable over time. EIERs are also geographic areas, and are roughly the same size and number as ERs, but they do not share the same definitions. Labour force estimates are produced for the EIERs for the use of Human Resources and Social Development Canada.

The intersections of the two types of regions form the first level of stratification for the LFS. These ER/EIER intersections are treated as primary strata and further stratification is carried out within them (see Section 5.2.3). Note that a third set of regions, census metropolitan areas (CMA), is also respected by stratification in the current LFS design, since each CMA is also an EIER.

Types of Areas
The primary strata (ER/EIER intersections) are further disaggregated into three types of areas: rural, urban and remote areas. Urban and rural areas are loosely based on the Census definitions of urban and rural, with some exceptions to allow for the formation of strata in some areas. Urban areas include the largest CMAs down to the smallest

[1] A detailed description of the LFS design is available in the Statistics Canada publication entitled *Methodology of the Canadian Labour Force Survey*, Catalogue no. 71-526-XPB.

[2] A detailed description of the CCHS design is available in the Statistics Canada publication entitled *Canadian Community Health Survey—Methodological overview*, Catalogue no. 82-003-XIE.

villages categorized by the 1991 Census as urban (1,000 people or more), while rural areas are made up of areas not designated as urban or remote.

All urban areas are further subdivided into two types: those using an apartment list frame and an area frame, as well as those using only an area frame.

Approximately 1% of the LFS population is found in remote areas of provinces, which are less accessible to LFS interviewers than other areas. For administrative purposes, this portion of the population is sampled separately through the remote area frame. Some populations, not congregated in places of 25 or more people, are excluded from the sampling frame.

Secondary Stratification
In urban areas with sufficiently large numbers of apartment buildings, the strata are subdivided into apartment frames and area frames. The apartment list frame is a register maintained for the 18 largest cities across Canada. The purpose of this is to ensure better representation of apartment dwellers in the sample as well as to minimize the effect of growth in clusters, due to construction of new apartment buildings. In the major cities, the apartment strata are further stratified into low income strata and regular strata.

Where it is possible and/or necessary, the urban area frame is further stratified into regular strata, high income strata, and low population density strata. Most urban areas fall into the regular urban strata, which, in fact, cover the majority of Canada's population. High income strata are found in major urban areas, while low density urban strata consist of small towns that are geographically scattered.

In rural areas, the population density can vary greatly from relatively high population density areas to low population density areas, resulting in the formation of strata that reflect these variations. The different stratification strategies for rural areas were based not only on concentration of population, but also on cost-efficiency and interviewer constraints.

In each province, remote settlements are sampled proportional to the number of dwellings in the settlement, with no further stratification taking place. Dwellings are selected using systematic random sampling in each of the places sampled.

Cluster Delineation and Selection
Households in the final strata are not selected directly. Instead, each stratum is divided into clusters, and then a sample of clusters is selected within the stratum. Dwellings are then sampled from selected clusters. Different methods are used to define the clusters, depending on the type of stratum.

Within each urban stratum in the urban area frame, a number of geographically contiguous groups of dwellings, or clusters, are formed based upon 1991 Census counts. These clusters are generally a set of one or more city blocks or block-faces. The selection of a sample of clusters (always six or a multiple of six clusters) from each of these secondary strata represents the first stage of sampling in most urban areas. In some other urban areas, census enumeration areas (EA) are used as clusters. In the low density urban strata, a three stage design is followed. Under this design, two towns within a stratum are sampled, and then 6 or 24 clusters within each town are sampled.

For urban apartment strata, instead of defining clusters, the apartment building is the primary sampling unit (PSU). Apartment buildings are sampled from the list frame with probability proportional to the number of units in each building.

Within each of the secondary strata in rural areas, where necessary, further stratification is carried out in order to reflect the differences among a number of socio-economic characteristics within each stratum. Within each rural stratum, six EAs or two or three groups of EAs are sampled as clusters.

Dwelling Selection
In all three types of areas (urban, rural and remote areas) selected clusters are first visited by enumerators in the field and a listing of all private dwellings in the cluster is prepared. From the listing, a sample of dwellings is then selected. The sample yield depends on the type of stratum. For example, in the urban area frame, sample yields are either six or eight dwellings, depending on the size of the city. In the urban apartment frame, each cluster yields five dwellings, while in the rural areas and EA parts of cities, each cluster yields 10 dwellings. In all clusters, dwellings are sampled systematically. This represents the final stage of sampling.

Person Selection
Demographic information is obtained for all persons in a household for whom the selected dwelling is the usual place of residence. LFS information is obtained for all civilian household members 15 years of age or older. Respondent burden is minimized for the elderly (age 70 and over) by carrying forward their responses for the initial interview to the subsequent five months in the survey.

5.1.3 Sample Size

The sample size of eligible persons in the LFS is determined so as to meet the statistical precision requirements for various labour force characteristics at the provincial and sub-provincial level, to meet the requirement of federal, provincial and municipal governments as well as a host of other data users.

The monthly LFS sample consists of approximately 60,000 dwellings. After excluding dwellings found to be vacant, dwellings demolished or converted to non-residential uses, dwellings containing only ineligible persons, dwellings under construction, and seasonal dwellings, about 54,000 dwellings remain which are occupied by one or more eligible persons. From these dwellings, LFS information is obtained from approximately 102,000 civilians aged 15 or over.

5.1.4 Sample Rotation

The LFS follows a rotating panel sample design, in which households remain in the sample for six consecutive months. The total sample consists of six representative sub-samples or panels, and each month a panel is replaced after completing its six month stay in the survey. Outgoing households are replaced by households in the same or a similar area. This results in a five-sixths month-to-month sample overlap, which makes the design efficient for estimating month-to-month changes. The rotation after six months prevents undue respondent burden for households that are selected for the survey.

Because of the rotation group feature, it is possible to readily conduct supplementary surveys using the LFS design but employing less than the full size sample.

5.1.5 Modifications to the Labour Force Survey Design for the Survey of Canadian Attitudes toward Learning

The SCAL used three rotation groups from the LFS sample that were last interviewed in October, November and December 2005. For the SCAL, coverage of the LFS was set at the household level. However, unlike the LFS where information is collected for all eligible household members, the SCAL only collected information from one pre-selected household member and proxy responses were not permitted.

Further, households consisting solely of persons less than 20 years of age were excluded from the SCAL. The table in Section 5.3 presents the sample sizes by province and territory.

5.2 Sample Selected from the Canadian Community Health Survey

The sampling frame for the Canadian Community Health Survey was used to select the sample for the Survey of Canadian Attitudes toward Learning in the three territories. The CCHS is a cross-sectional survey that collects information on the state of health, the use of health services and determinants of health for the Canadian population. The collection cycle for the CCHS extended over two years. The first year of the cycle, identified by the suffix ".1", constitutes a general survey on the health of the population with a large sample designed to provide reliable estimate at the health region level. The second year of the cycle, identified by the suffix ".2", consists of a smaller sample and is designed to provide reliable estimates at the province level on special topics related to health.

5.2.1 Target Population

The CCHS targets persons aged 12 years and older who are living in private dwellings in the 10 provinces and three territories. Persons living on Indian Reserves or Crown lands, those residing in institutions, full-time members of the Canadian Armed Forces and residents of certain remote regions are excluded from this survey. The CCHS covered approximately 98% of the Canadian population aged 12 and older.

Further, households consisting solely of persons less than 20 years of age were excluded from the SCAL. The table in Section 5.3 presents the sample sizes by province and territory.

5.2.2 Health Regions

For administrative purposes, each province is divided into health regions (HR) and each territory is designated as a single HR (see table below). Statistics Canada, in consultation with the provinces, made minor changes to the boundaries of some of the HRs to correspond to the geography of the 2001 Census. During Cycle 3.1 of the CCHS, data was collected in 122 HRs in the 10 provinces, in addition to one HR per territory, totalling 125 HRs.

Number of health regions and targeted sample sizes by territory

Territories	Number of Health Regions	Total targeted sample size
Yukon	1	850
Northwest Territories	1	900
Nunavut	1	700
Total	**3**	**2,450**

5.2.3 Sample Design

Cycle 3.1 of the CCHS used three sampling frames to select the sample of households: 49% of the sample of households came from an area frame, 50% came from a list frame of telephone numbers and the remaining 1% came from a Random Digit Dialling (RDD) sampling frame. For the SCAL, only households from the area frame were used.

Sampling of Households from the Area Frame
The CCHS used the area frame designed for the Labour Force Survey as a sampling frame. Requirements specific to the CCHS led to some modifications to the LFS sampling strategy. To get a base sample of 62,000 households for the CCHS, 86,000 dwellings must be selected from the area frame (to account for vacant dwellings and non-responding households). On an on-going monthly basis the LFS design provides approximately 60,000 dwellings distributed across the various economic regions in Canada whereas the CCHS requires a total of 86,000 dwellings distributed in the HRs, which have different geographic boundaries than those of the LFS economic regions. Overall, the CCHS required 43% more dwellings than those generated by the LFS selection mechanism, for an *adjustment factor* of 1.43 (86,000 / 60,000). However, at the HR level, the adjustment factors varied from 0.6 to 6.0.

The changes made to the selection mechanism in a HR varied depending on the size of the adjustment factors. For HRs that had a factor less than or equal to 1, a simple stabilization, was applied to the sample of dwellings. For those with a factor greater than 1 but less than or equal to 2, the sampling process of dwellings within a PSU was repeated for all selected PSUs that were part of the same HR. For HRs with a factor greater than 2 but less than or equal to 4, the PSU sampling process, as well as that for dwellings in a PSU, was repeated. For HRs with a factor between 4 and 6, the PSU sampling process was repeated not once but twice while that for dwellings was repeated only once. Where the chosen approach created an unnecessary surplus of dwellings, stabilization was performed.

It should be noted that the changes made to the LFS mechanism resulted in, at most, tripling the number of PSUs selected and, at most, doubling the number of dwellings selected in the PSUs, which explained the maximum adjustment factor of 6.0. At the HR level, adjustment factors were purposely capped at 6.0 for two reasons: to limit the listing of clusters (each newly selected PSU requires a listing), and to avoid possible cluster effects created by too great a number of dwellings selected in a single PSU. This limit to the adjustment factor of certain HRs has consequently dictated the number of households required from the telephone frames.

Sampling of Households from the Area Frame in the Three Territories
For operational reasons the area frame sample design implemented in the three northern territories had one additional stage of selection. For each territory, in-scope communities were first stratified based on various characteristics (population, geography, proportion of Inuit and/or Aboriginal persons, and median household income). There were five design strata in the Yukon, ten in the Northwest Territories and six in Nunavut. The first stage of selection consisted of randomly selecting one community with a probability proportional to population size within each design stratum. From that point on, the household sampling strategy from the area frame within the selected community was identical to the one described above.

It is worth mentioning that the frame for the CCHS covered 90% of the private households in the Yukon, 97% in the Northwest Territories and 71% in Nunavut.

5.2.4 Modifications to the Canadian Community Health Survey Design for the Survey of Canadian Attitudes toward Learning

The frame used to select the sample in the territories contained 2,658 households. The sample for the territories was different than in the provinces, where only households solely consisting of members less than 20 years of age were excluded. In fact, the 181 households from the RDD (random digit dialling) frame of the CCHS were excluded in order to simplify the design. Further, the 277 households for which telephone numbers were not available were excluded in order to avoid selecting households that do not have telephone service - a situation that is more common in the North than elsewhere in Canada. Finally, two households were excluded because all members were less than 20 years of age.

Similarly to the provinces, the number of households selected in each territory was pre-determined (193 in Yukon, 251 in Northwest Territories and 180 in Nunavut). The households were therefore selected using simple random sampling in each territory, and then one member in each household was selected using the same method.

From the 2,198 remaining eligible households 624 households were selected.

Sample sizes in the territories for the CCHS, Cycle 3.1 and the SCAL 2006

Territories	Number of Health Regions	Households in-scope	Households responding to the CCHS	Households used to select the sample for the SCAL	Households selected in the SCAL sample
Yukon	1	926	868	666	193
Northwest Territories	1	1,121	1,007	881	251
Nunavut	1	893	783	651	180
Total	**3**	**2,940**	**2,658**	**2,198**	**624**

5.3 Sample Size by Province/Territory for the Survey of Canadian Attitudes toward Learning

The following table shows the number of household members in the LFS sampled rotations who were eligible for the Survey of Canadian Attitudes toward Learning. This table includes households which were non-respondents to the LFS and the CCHS.

Provinces and Territories	Sample Size
Provinces	**7,946**
Newfoundland and Labrador	515
Prince Edward Island	504
Nova Scotia	527
New Brunswick	523
Quebec	1,572
Ontario	1,270
Manitoba	572
Saskatchewan	527
Alberta	746
British Columbia	1,190
Territories	**624**
Yukon	193
Northwest Territories	251
Nunavut	180
Canada	**8,570**

6.0 Data Collection

Data collection for the Survey of Canadian Attitudes toward Learning (SCAL) was carried out between April 25th and May 30th, 2006. The SCAL was administered to one randomly selected individual per household using a computer-assisted telephone interview (CATI) questionnaire. The random selection was carried out prior to the beginning of data collection during the sample selection (see Section 5.1.5 and 5.2.4).

Since the data collection period for the SCAL followed by several months the last interview for households selected from the Labour Force Survey (LFS) in the provinces and from the Canadian Community Health Survey (CCHS) in the territories, interviewers proceeded to a tracing exercise when the selected respondent for the SCAL had moved or when the telephone number on our files was not valid or out of service.

6.1 Supervision and Quality Control

All interviewers are under the supervision of a staff of senior interviewers who are responsible for ensuring that interviewers are familiar with the concepts and procedures of the survey. They are also responsible for periodically monitoring their interviewers and reviewing their completed documents. The senior interviewers are, in turn, under the supervision of program managers, located in each of the Statistics Canada regional offices.

6.2 Non-response to the Survey of Canadian Attitudes toward Learning

Interviewers are instructed to make all reasonable attempts to obtain interviews with the selected respondent. For individuals who at first refuse to participate, a letter is sent from the Regional Office to the dwelling address stressing the importance of the survey and the household's cooperation. This is followed by a second call from the interviewer. For cases in which the timing of the interviewer's call is inconvenient, an appointment is arranged to call back at a more convenient time. For cases in which there is no one home, numerous call backs are made. Under no circumstances are sampled respondents replaced by other individuals for reasons of non-response.

7.0 Data Processing

The main output of the Survey of Canadian Attitudes toward Learning (SCAL) is a "clean" microdata file. This chapter presents a brief summary of the processing steps involved in producing this file.

7.1 Data Capture

Responses to survey questions are captured directly by the interviewer at the time of the interview using a computerized questionnaire. The computerized questionnaire reduces processing time and costs associated with data entry, transcription errors and data transmission. The response data are encrypted to ensure confidentiality and sent via modem to the appropriate Statistics Canada Regional Office. From there they are transmitted over a secure line to Ottawa for further processing.

Some editing is done directly at the time of the interview. Where the information entered is out of range (too large or small) of expected values, or inconsistent with the previous entries, the interviewer is prompted, through message screens on the computer, to modify the information. However, for some questions interviewers have the option of bypassing the edits, and of skipping questions if the respondent does not know the answer or refuses to answer. Therefore, the response data are subjected to further edit and imputation processes once they arrive in head office.

7.2 Editing

The first stage of survey processing undertaken at head office was the replacement of any "out-of-range" values on the data file with blanks. This process was designed to make further editing easier.

The first type of error treated was errors in questionnaire flow, where questions which did not apply to the respondent (and should therefore not have been answered) were found to contain answers. In this case a computer edit automatically eliminated superfluous data by following the flow of the questionnaire implied by answers to previous, and in some cases, subsequent questions.

The second type of error treated involved a lack of information in questions which should have been answered. For this type of error, a non-response or "not-stated" code was assigned to the item.

7.3 Coding of Open-ended Questions

A few data items on the questionnaire were recorded by interviewers in an open-ended format. A total of two partially or completely open-ended questions were included in the survey. These were items relating to country of birth and ethnic origin.

7.4 Imputation

Imputation is the process that supplies valid values for those variables that have been identified for a change either because of invalid information or because of missing information. The new values are supplied in such a way as to preserve the underlying structure of the data and to ensure that the resulting records will pass all required edits. In other words, the objective is not to reproduce the true microdata values, but rather to establish internally consistent data records that yield good aggregate estimates.

We can distinguish between three types of non-response. Complete non-response is when the respondent does not provide the minimum set of answers. These records are dropped and accounted for in the weighting process (see Chapter 11.0). Item non-response is when the respondent does not provide an answer to one question, but goes on to the next question. These are usually handled using the "not stated" code or are imputed. Finally, partial non-response is when the respondent provides the minimum set of answers but does not finish the interview. These records can be handled like either complete non-response or multiple item non-response.

Further information on the imputation process is given in Chapter 8.0 (Data Quality).

7.5 Creation of Derived Variables

A number of data items on the microdata file have been derived by combining items on the questionnaire in order to facilitate data analysis. These variables are identified in the codebook. An important derived variable (DV) is LFSTAT which determines the labour force status of the respondent from the responses that were given in the Labour Force (LF) module. Other examples include a derived variable to determine the highest level of education (EDUCLEV) reached. Various other DVs are created to categorize continuous variables and to facilitate the information collected on the respondent's children (age and gender).

7.6 Weighting

The principle behind estimation in a probability sample such as the LFS is that each person in the sample "represents", besides himself or herself, several other persons not in the sample. For example, in a simple random 2% sample of the population, each person in the sample represents 50 persons in the population.

The weighting phase is a step which calculates, for each record, what this number is. This weight appears on the microdata file, and **must** be used to derive meaningful estimates from the survey. For example if the number of individuals who have taken a program or training related to work during the past 12 months is to be estimated, it is done by selecting the records referring to those individuals in the sample with that characteristic and summing the weights entered on those records.

Details of the method used to calculate these weights are presented in Chapter 11.0.

8.0 Data Quality

8.1 Response Rates

The following table summarizes the response rates to the Labour Force Survey (LFS), the Canadian Community Health Survey (CCHS) and to the Survey of Canadian Attitudes toward Learning (SCAL).

Provinces and Territories	LFS in the Provinces CCHS in the Territories*			SCAL**			Overall response rate to SCAL (%)***
	Selected households	Responding households	Response rate (%)	Selected households	Responding households	Response rate (%)	
	(1)	(2)	(3)	(4)	(5)	(6)	(3)*(6)
Newfoundland and Labrador	938	861	91.8	515	314	61.0	56.0
Prince Edward Island	727	673	92.6	504	315	62.5	57.9
Nova Scotia	1,434	1,333	93.0	527	324	61.5	57.1
New Brunswick	1,347	1,255	93.2	523	351	67.1	62.5
Quebec	4,915	4,471	91.0	1,572	950	60.4	55.0
Ontario	7,699	7,094	92.1	1,270	763	60.1	55.4
Manitoba	1,803	1,682	93.3	572	345	60.3	56.3
Saskatchewan	1,892	1,755	92.8	527	334	63.4	58.8
Alberta	2,687	2,468	91.8	746	475	63.7	58.5
British Columbia	3,146	2,904	92.3	1,190	687	57.7	53.3
Yukon	926	827	89.3	193	143	74.1	66.2
Northwest Territories	1,121	1,007	89.8	251	165	65.7	59.1
Nunavut	893	830	92.9	180	100	55.6	51.6
Canada	**29,528**	**27,160**	**92.0**	**8,570**	**5,266**	**61.4**	**56.5**

Note: The LFS and CCHS counts represent households while the SCAL counts represent selected individuals within households (only one individual is selected per household). The SCAL overall response rate is based on all LFS and CCHS household records. The LFS responding households include respondents carried forward from the previous month.

* The LFS response rate is the number of LFS responding households as a percentage of the number of LFS selected households.

** The SCAL response rate is the number of SCAL responding households who agreed to share their responses as a percentage of the number of households selected for the SCAL.

*** The SCAL overall response rate is the number of responding persons who accepted to share their responses during the SCAL interview as a percentage of the number of LFS and CCHS selected households.

8.2 Survey Errors

The estimates derived from this survey are based on a sample of households. Somewhat different estimates might have been obtained if a complete census had been taken using the same questionnaire, interviewers, supervisors, processing methods, etc. as those actually used in the survey. The difference between the estimates obtained from the sample and those resulting 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.

Over a large number of observations, randomly occurring 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 were taken 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 include 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.

8.2.1 The Frame

Given that the SCAL was a supplement to the LFS and the CCHS, the sampling frame is made up of households that responded to one of these surveys. Any non-response to the LFS had an impact on the SCAL frame. The quality of the sampling variables in the frame was very high. The SCAL sample consisted of three rotation groups from the LFS. The criteria used for the SCAL selection were not missing for any LFS or CCHS records.

Note that the LFS frame excludes about 2% of all households in the 10 provinces of Canada. Therefore, the SCAL frame also excludes the same proportion of households in the same geographical area. It is unlikely that this exclusion introduces any significant bias into the survey data.

The computer-assisted personal interview (CAPI) cases selected from the LFS were automatically set to “out-of-scope” without any to trace even though an address was available. Tracing is used in order to establish contact for LFS cases that were conducted as computer-assisted telephone interviews (CATI) when the telephone number was no longer valid or out of service.

The 2,658 households included in the area frame based sample for the CCHS constitute the frame for the SCAL in the territories. Of these households, 277 cases were excluded in order to exclude cases that effectively do not have access to telephone service: a situation that is quite common in the North. The frame also excluded a two households that did not contain any members in the target population (i.e. 20 years of age and older).

8.2.2 Data Collection

Interviewer training consisted of reading the SCAL Procedures Manual and the Interviewer’s Manual, practicing with the SCAL training cases on the computer, and discussing any questions with senior interviewers before the start of the survey. A description of the background and objectives of the survey was provided, as well as a glossary of terms and a set of questions and answers. The collection period ran from the week of April 25th to May 30th, 2006.

8.2.3 Data Processing

During processing of the data, some records were discarded because respondents did not agree to share their responses with the Canadian Council on Learning. There were 206 such records and these were coded as non-response.

Data processing of the SCAL was done in a number of steps including verification, coding, editing, imputation, estimation, confidentiality, etc. At each step a picture of the output files is taken and an easy verification can be made by comparing files at the current and previous step. This greatly improved the data processing stage.

8.2.4 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 because 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. Total non-response was handled by adjusting the weight of individuals who responded to the survey to compensate for those who did not respond.

In most cases, partial non-response to the survey occurred when the respondent did not understand or misinterpreted a question, refused to answer a question, or could not recall the requested information.

8.2.5 Measurement of Sampling Error

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 (CV) 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 18.65% of persons 20 years of age and older have at least one child aged between 0 and 8 years and this estimate is found to have a standard error of 0.0135. Then the coefficient of variation of the estimate is calculated as:

$$\left(\frac{0.0135}{0.1865}\right) X\ 100\,\% = 7.24\,\%$$

There is more information on the calculation of coefficients of variation in Chapter 10.0.

9.0 Guidelines for Tabulation, Analysis and Release

This chapter of the documentation outlines the guidelines to be adhered to by users tabulating, analyzing, publishing or otherwise releasing any data derived from the survey microdata files. 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.

9.1 Rounding Guidelines

In order that estimates for publication or other release derived from these microdata files 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 subtotals 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.

9.2 Sample Weighting Guidelines for Tabulation

The sample design used for the Survey of Canadian Attitudes toward Learning (SCAL) was not self-weighting. When producing simple estimates including the production of ordinary statistical tables, users must apply the proper survey weights.

If proper weights are not used, the estimates derived from the microdata files cannot be considered to be representative of the survey population, and will not correspond to those produced by Statistics Canada.

Users should also note that some software packages may not allow the generation of estimates that exactly match those available from Statistics Canada, because of their treatment of the weight field.

9.3 Definitions of Types of Estimates: Categorical and Quantitative

Before discussing how the SCAL 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 SCAL.

9.3.1 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 persons who participated in work-related training during the last year and the number of persons who have at least one child aged between 0 and 8 years who has been in any regular day care arrangement including at a day care centre in the last five years 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: During the last year, did you take a program towards a degree, diploma or certificate or take any courses, workshops, seminars or formal training for any reasons related to a current or future job?
A: Yes / No

Q: In the past 5 years, have any of your children…?
A: Been in any regular day care arrangement including at a day care centre / Been under the care of a caregiver in someone else's home / Been under the care of a caregiver other than a relative inside your own home / None of the above

9.3.2 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 $\hat{X}$ 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 children that persons 20 years of age and older have. The numerator is an estimate of the total number of children that persons 20 years of age and older have and its denominator is the number of persons 20 years of age and older who have at least one child.

Examples of Quantitative Questions:

Q: How many children do you have? Please include children that you are the legal guardian of and also include children that do not live with you.
A: |_|_| children

Q: How much income did all members of your household receive in total from all the sources mentioned before taxes and deductions, in the year ending December 31, 2005?
A: |_|_|_|_|_|_|_|_| dollars

9.3.3 Tabulation of Categorical Estimates

Estimates of the number of people 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 $\hat{X} / \hat{Y}$ are obtained by:

a) summing the final weights of records having the characteristic of interest for the numerator ($\hat{X}$),
b) summing the final weights of records having the characteristic of interest for the denominator ($\hat{Y}$), then
c) dividing estimate a) by estimate b) ($\hat{X} / \hat{Y}$).

9.3.4 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 of the total number of children that persons 20 years of age and older have, multiply the value reported in question SQ_Q02 (number of children that the respondent has, whether they live with him or not) by the final weight for the record, then sum this value over all records with SQ_Q01 = 1 (have children).

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 children that persons 20 years of age and older have.

a) estimate the total number of children ($\hat{X}$) as described above,
b) estimate the total number of persons 20 years of age and older who have children ($\hat{Y}$) in this category by summing the final weights of all records with SQ_Q01 = 1, then
c) divide estimate a) by estimate b) ($\hat{X} / \hat{Y}$).

9.4 Guidelines for Statistical Analysis

The SCAL is based upon a complex sample design, 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 may 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. Approximate variances for simple estimates such as totals, proportions and ratios (for qualitative variables) can be derived using the accompanying Approximate Sampling Variability Tables.

For other analysis techniques (for example linear regression, logistic regression and analysis of variance), a method exists which can make the variances calculated by the standard packages more meaningful, by incorporating the unequal probabilities of selection. The method rescales the weights so that there is an average weight of 1.

For example, suppose that analysis of all male respondents is required. The steps to rescale the weights are as follows:

1) select all respondents from the file who reported LF_Q01 = men;

2) calculate the AVERAGE weight for these records by summing the original person weights from the microdata file for these records and then dividing by the number of respondents who reported LF_Q01 = men;

3) for each of these respondents, calculate a RESCALED weight equal to the original person weight divided by the AVERAGE weight;

4) perform the analysis for these respondents using the RESCALED weight.

However, because the stratification and clustering of the sample's design are still not taken into account, the variance estimates calculated in this way are likely to be under-estimates.

The calculation of more precise variance estimates requires detailed knowledge of the design of the survey. Such detail cannot be given in this microdata file because of confidentiality. Variances that take the complete sample design into account can be calculated for many statistics by Statistics Canada on a cost-recovery basis.

9.5 Coefficient of Variation Release Guidelines

Before releasing and/or publishing any estimates from the SCAL, 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 as discussed in Chapter 8.0. However for this purpose, 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 the table below. Nonetheless users should be sure to read Chapter 8.0 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 rounded weighted 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.

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 of 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 of 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 releasing 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: "Please be warned that these estimates [flagged with the letter U] do not meet Statistics Canada's quality standards. Conclusions based on these data will be unreliable, and most likely invalid."

9.6 Release Cut-offs for the Survey of Canadian Attitudes toward Learning

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 the previous section. These cut-offs are derived from the coefficient of variation (CV) tables discussed in Chapter 10.0.

For example, the table shows that the quality of a weighted estimate of 18,500 people possessing a given characteristic in Newfoundland and Labrador is marginal.

Note that these cut-offs apply to estimates of population totals only. To estimate ratios, users should not use the numerator value (or the denominator) in order to find the corresponding quality level. Rule 4 in Section 10.1 and Example 4 in Section 10.1.1 explain the correct procedure to be used for ratios.

Provinces, Territories and Regions of Residence in 2006	Acceptable CV 0.0% to 16.5%		Marginal CV 16.6% to 33.3%			Unacceptable CV > 33.3%	
Atlantic Provinces	95,500	& over	24,500	to <	95,500	under	24,500
Newfoundland and Labrador	66,000	& over	18,500	to <	66,000	under	18,500
Prince Edward Island	15,000	& over	4,000	to <	15,000	under	4,000
Nova Scotia	112,000	& over	31,500	to <	112,000	under	31,500
New Brunswick	81,500	& over	22,500	to <	81,500	under	22,500
Québec	485,000	& over	127,000	to <	485,000	under	127,000
Ontario	1,436,500	& over	399,000	to <	1,436,500	under	399,000
Prairie Provinces	294,000	& over	76,500	to <	294,000	under	76,500
Manitoba	132,000	& over	37,000	to <	132,000	under	37,000
Saskatchewan	105,500	& over	29,500	to <	105,500	under	29,500
Alberta	365,500	& over	101,500	to <	365,500	under	101,500
British Colombia	347,000	& over	92,500	to <	347,000	under	92,500
Territories	8,000	& over	2,000	to <	8,000	under	2,000
Yukon	6,000	& over	2,000	to <	6,000	under	2,000
Northwest Territories	7,500	& over	2,000	to <	7,500	under	2,000
Nunavut	6,500	& over	2,000	to <	6,500	under	2,000
Canada	**873,000**	**& over**	**220,500**	**to <**	**873,000**	**under**	**220,500**

10.0 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 Survey of Canadian Attitudes toward Learning (SCAL).

Provinces, Territories and Regions	Design Effect	Sample Size	Population
Atlantic Provinces	2.04	1,304	1,760,416
Newfoundland and Labrador	1.72	314	394,362
Prince Edward Island	1.50	315	101,973
Nova Scotia	1.68	324	700,417
New Brunswick	1.62	351	563,664
Quebec	2.37	950	5,775,934
Ontario	3.76	763	9,372,303
Prairie Provinces	2.57	1,154	3,886,972
Manitoba	1.82	345	813,610
Saskatchewan	1.68	334	677,737
Alberta	2.33	475	2,395,625
British Columbia	2.25	687	3,230,139
Territories	1.50	408	68,426
Yukon	1.42	143	23,356
Northwest Territories	1.54	165	28,482
Nunavut	1.65	100	16,588
Canada	**5.39**	**5,266**	**24,094,190**

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. Since the approximate CV is 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 releasing such an estimate, regardless of the value of the coefficient of variation.

10.1 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 subgroup of the population. For example, the proportion of persons who have at least one child aged between 0 and 8 years is more reliable than the estimated number of persons who have at least one child aged between 0 and 8 years. (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 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 persons who have at least one child aged between 0 and 8 years and the numerator is the number of persons who have at least one child aged between 0 and 8 years who has been in any regular day care arrangement including at a day care centre in the last five years.

In the case where the numerator is not a subset of the denominator, as for example, the ratio of the number of men who have at least one child aged between 0 and 8 years as compared to the number of women who have at least one child aged between 0 and 8 years, 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.

10.1.1 Examples of Using the Coefficient of Variation Tables for Categorical Estimates

The following examples based on the SCAL 2006 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 4,493,502 persons have at least one child aged between 0 and 8 years. How does the user determine the coefficient of variation of this estimate?

1) Refer to the coefficient of variation table for CANADA.

2) The estimated aggregate (4,493,502) 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 4,000,000.

3) The coefficient of variation for an estimated aggregate is found by referring to the first non-asterisk entry on that row, namely, 7.0%.

4) So the approximate coefficient of variation of the estimate is 7.0%. The finding that there were 4,493,502 (to be rounded according to the rounding guidelines in Section

9.1) persons have at least one child aged between 0 and 8 years is publishable with no qualifications.

Survey of Canadian Attitudes Toward Learning, 2006

Approximate Sampling Variability Tables - Canada

NUMERATOR OF PERCENTAG ('000)	ESTIMATED PERCENTAGE												
	0.1%	**1.0%**	**2.0%**	**5.0%**	**10.0%**	**15.00%**	**20.0%**	**...**	**35.0%**	**40.0%**	**50.0%**	**70.0%**	**90.0%**
1	496.3	494.1	491.6	484.0	471.1	457.8	444.1	...	400.3	384.6	351.1	272.0	157.0
2	350.9	349.4	347.6	342.2	333.1	323.7	314	...	283.1	272.0	248.3	192.3	111.0
3	286.5	285.2	283.8	279.4	272.0	264.3	256.4	...	231.1	222.1	202.7	157.0	90.7
4	248.2	247.0	245.8	242.0	235.5	228.9	222.1	...	200.2	192.3	175.6	136.0	78.5
5	222.0	221.0	219.8	216.4	210.7	204.7	198.6	...	179.0	172.0	157.0	121.6	70.2
6	202.6	201.7	200.7	197.6	192.3	186.9	181.3	...	163.4	157.0	143.3	111.0	64.1
7	187.6	186.7	185.8	182.9	178.0	173	167.9	...	151.3	145.4	132.7	102.8	59.3
8	175.5	174.7	173.8	171.1	166.5	161.9	157	...	141.5	136.0	124.1	96.2	55.5
9	165.4	164.7	163.9	161.3	157.0	152.6	148	...	133.4	128.2	117.0	90.7	52.3
10	156.9	156.2	155.4	153.0	149.0	144.8	140.4	...	126.6	121.6	111.0	86.0	49.7
11	149.6	149.0	148.2	145.9	142.0	138	133.9	...	120.7	116.0	105.9	82.0	47.3
12	143.3	142.6	141.9	139.7	136.0	132.2	128.2	...	115.6	111.0	101.4	78.5	45.3
.	.	.	.	.	.	.	.	...	.	.	.	.	.
.	.	.	.	.	.	.	.	...	.	.	.	.	.
.	.	.	.	.	.	.	.	...	.	.	.	.	.
450	*****	*****	23.2	22.8	22.2	21.6	20.9	...	18.9	18.1	16.6	12.8	7.4
500	*****	*****	*****	21.6	21.1	20.5	19.9	...	17.9	17.2	15.7	12.2	7.0
750	*****	*****	*****	**17.7**	17.2	16.7	16.2	...	14.6	**14.0**	12.8	9.9	5.7
1,000	*****	*****	*****	**15.3**	14.9	14.5	14.0	...	12.7	12.2	**11.1**	8.6	5.0
1,500	*****	*****	*****	*****	12.2	11.8	11.5	...	10.3	9.9	9.1	7.0	4.1
2,000	******	*****	*****	*****	10.5	10.2	9.9	...	9.0	**8.6**	7.9	6.1	3.5
3,000	*****	*****	*****	*****	*****	8.4	8.1	...	7.3	7.0	6.4	5.0	2.9
4,000	*****	*****	*****	*****	*****	*****	**7.0**	...	6.3	6.1	5.6	4.3	2.5
5,000	*****	*****	*****	*****	*****	*****	*****	...	5.7	5.4	5.0	3.8	2.2
6,000	*****	*****	*****	*****	*****	*****	*****	...	5.2	5.0	4.5	3.5	2.0
7,000	*****	*****	*****	*****	*****	*****	*****	...	4.8	4.6	4.2	3.3	1.9
8,000	*****	*****	*****	*****	*****	*****	*****	...	4.5	4.3	3.9	3.0	1.8
9,000	*****	*****	*****	*****	*****	*****	*****	...	*****	4.1	3.7	2.9	1.7
10,000	*****	*****	*****	*****	*****	*****	*****	...	*****	*****	3.5	2.7	1.6
12,500	*****	*****	*****	*****	*****	*****	*****	...	*****	*****	*****	2.4	1.4
15,000	*****	*****	*****	*****	*****	*****	*****	...	*****	*****	*****	2.2	1.3
20,000	*****	*****	*****	*****	*****	*****	*****	...	*****	*****	*****	*****	1.1

NOTE: For correct usage of these tables, please refer to the microdata documentation.

Example 2: Estimates of Proportions or Percentages of Persons Possessing a Characteristic

Suppose that the user estimates that 1,920,409 / 4,493,502 = 42.7% of persons have at least one child aged between 0 and 8 years who has been in any regular day care arrangement including at a day care centre in the last five years. How does the user determine the coefficient of variation of this estimate?

1) Refer to the coefficient of variation table for CANADA.

2) Because the estimate is a percentage which is based on a subset of the total population (i.e., persons who have at least one child aged between 0 and 8 years), it

is necessary to use both the percentage (42.7%) and the numerator portion of the percentage (1,920,409) in determining the coefficient of variation.

3) The numerator, 1,920,409, 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 2,000,000. 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 8.6% is the coefficient of variation to be used.

5) So the approximate coefficient of variation of the estimate is 8.6%. The finding that 42.7% of persons have at least one child aged between 0 and 8 years who has been in any regular day care arrangement including at a day care centre in the last five years can be published with no qualifications.

Example 3: Estimates of Differences between Aggregates or Percentages

Suppose that a user estimates that 775,148 / 1,992,794 = 38.9% of women have at least one child aged between 0 and 8 years who has been in any regular day care arrangement including at a day care centre in the last five years, while 1,145,261 / 2,500,708 = 45.8% of men have at least one child aged between 0 and 8 years who has been in any regular day care arrangement including at a day care centre in the last five years. 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 14.0%, and the CV of the estimate for men as 11.1%.

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 $\hat{d}$ = 0.389 – 0.458 = -0.069 is:

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(0.389)(0.14)]^2 + [(0.458)(0.111)]^2} \\ &= \sqrt{(0.0029658) + (0.0025845)} \\ &= 0.0745\end{aligned}$$

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d}$ = 0.0745 / 0.069 = 1.0797

4) So the approximate coefficient of variation of the difference between the estimates is 108.0%. 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 775,148 women have at least one child aged between 0 and 8 years who has been in any regular day care arrangement including at a day care centre in the last five years, while 1,145,261 men have at least one child aged between 0 and 8 years who has been in any regular day care arrangement including at a day care centre in the last five years. The user is interested in comparing the estimate of women versus that of men 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 women who have at least one child aged between 0 and 8 years who has been in any regular day care arrangement including at a day care centre in the last five years. The denominator of the estimate ($\hat{X}_2$) is the number of men who have at least one child aged between 0 and 8 years who has been in any regular day care arrangement including at a day care centre in the last five years.

2) Refer to the coefficient of variation table for CANADA.

3) The numerator of this ratio estimate is 775,148. The figure closest to it is 750,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 17.7%.

4) The denominator of this ratio estimate is 1,145,261. The figure closest to it is 1,000,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 15.3%.

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:

$$\begin{aligned} \alpha_{\hat{R}} &= \sqrt{(0.177)^2 + (0.153)^2} \\ &= \sqrt{0.031329 + 0.023409} \\ &= 0.23396 \end{aligned}$$

6) The obtained ratio of women versus men who have at least one child aged between 0 and 8 years who has been in any regular day care arrangement including at a day care centre in the last five years, is 775,148 / 1,145,261 which is 0.677 (to be rounded according to the rounding guidelines in Section 9.1). The coefficient of variation of this estimate is 23.4%, which makes the estimate marginal even though it is releasable. This estimate 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 estimate.

Example 5: Estimates of Differences of Ratios

Suppose that the user estimates that the ratio of women who have at least one child aged between 0 and 8 years who has been in any regular day care arrangement including at a day care centre in the last five years and men who have at least one child aged between 0 and 8 years who has been in any regular day care arrangement including at a day care centre in the last five years, is 1.06 in Quebec, while it is 0.71 for the Atlantic Provinces. 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 Quebec ratio ($\hat{R}_1$) and the Atlantic Provinces ratio ($\hat{R}_2$) as in Example 4. The approximate CV for the Quebec ratio is 29.4% and 29.9% for the Atlantic Provinces.

2) Using Rule 3, the standard error of a difference ($\hat{d} = \hat{R}_1 - \hat{R}_2$) is:

$$\sigma_{\hat{d}} = \sqrt{\left(\hat{R}_1\alpha_1\right)^2 + \left(\hat{R}_2\alpha_2\right)^2}$$

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 $\hat{d}$ = 1.06 – 0.71 = 0.35 is:

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(1.06)(0.294)]^2 + [(0.71)(0.299)]^2} \\ &= \sqrt{(0.097) + (0.045)} \\ &= 0.377\end{aligned}$$

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d}$ = 0.377 / 0.35 = 1.077.

4) So the approximate coefficient of variation of the difference between the estimates is 107.7%. 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.

10.2 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.

10.2.1 Example of Using the Coefficient of Variation Tables to Obtain Confidence Limits

A 95% confidence interval for the estimated proportion of persons who have at least one child aged between 0 and 8 years who has been in any regular day care arrangement including at a day care centre in the last five years (from Example 2, Section 10.1.1) would be calculated as follows:

$\hat{X}$ = 42.7% (or expressed as a proportion 0.427)

t = 2

$\alpha_{\hat{x}}$ = 8.6% (0.086 expressed as a proportion) is the coefficient of variation of this estimate as determined from the tables.

$CI_{\hat{x}}$ = {0.427– (2) (0.427) (0.086), 0.427+ (2) (0.427) (0.086)}

$$CI_{\hat{x}} = \{0.427 - 0.073,\ 0.427 + 0.073\}$$

$$CI_{\hat{x}} = \{0.354,\ 0.500\}$$

With 95% confidence it can be said that between 35.4% and 50.0% of persons have at least one child aged between 0 and 8 years who has been in any regular day care arrangement including at a day care centre in the last five years.

10.3 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 $t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}}$ 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.

10.3.1 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 women who have at least one child aged between 0 and 8 years who has been in any regular day care arrangement including at a day care centre in the last five years and the proportion of men who have at least one child aged between 0 and 8 years who has been in any regular day care arrangement including at a day care centre in the last five years. From Example 3, Section 10.1.1, the standard error of the difference between these two estimates was found to be 0.0745. Hence,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0.389 - 0.458}{0.0745} = \frac{-0.069}{0.0745} = -0.926$$

Since t = -0.926 is between -2 and 2, it must be concluded that there is no significant difference between the two estimates at the 0.05 level of significance.

10.4 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 SCAL 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 number of children of persons who have at least one child aged between 0 and 8 years who has been in any regular day care arrangement including at a day care centre in the last five years would be greater than the coefficient of variation of the corresponding proportion of persons who have at least one child aged between 0 and 8 years who has been in any regular day care arrangement including at a day care centre in the last five years. 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 as required for a specific estimate using a technique known as pseudo replication. This involves dividing the records on the microdata files into subgroups (or replicates) and determining the variation in the estimate from replicate to replicate. Users wishing to derive coefficients of variation for quantitative estimates may contact Statistics Canada for advice on the allocation of records to appropriate replicates and the formulae to be used in these calculations.

10.5 Coefficient of Variation Tables

Refer to SCAL2006_CVTabsE.pdf for the coefficient of variation tables.

11.0 Weighting

Since the Survey of Canadian Attitudes toward Learning (SCAL) used a sub-sample of the Labour Force Survey (LFS) and the Canadian Community Health Survey (CCHS) sample, the derivation of weights for the survey records is clearly tied to the weighting procedure used for the LFS and the CCHS. The LFS and the CCHS weighting procedure are briefly described below.

11.1 Weighting Procedures for the Labour Force Survey

In the LFS, the final weight attached to each record is the product of the following factors: the basic weight, the cluster sub-weight, the stabilization weight, the balancing factor for non-response, and the province–age–sex and sub-provincial area ratio adjustment factor. Each is described below.

Basic Weight
In a probability sample, the sample design itself determines weights which must be used to produce unbiased estimates of population. Each record must be weighted by the inverse of the probability of selecting the person to whom the record refers. In the example of a 2% simple random sample, this probability would be 0.02 for each person and the records must be weighted by 1 / 0.02 = 50. Due to the complex LFS design, dwellings in different regions will have different basic weights. Because all eligible individuals in a dwelling are interviewed (directly or by proxy), this probability is essentially the same as the probability with which the dwelling is selected.

Cluster Sub-weight
The cluster delineation is such that the number of dwellings in the sample increases very slightly with moderate growth in the housing stock. Substantial growth can be tolerated in an isolated cluster before the additional sample represents a field collection problem. However, if growth takes place in more than one cluster in an interviewer assignment, the cumulative effect of all increases may create a workload problem. In clusters where substantial growth has taken place, sub-sampling is used as a means of keeping interviewer assignments manageable. The cluster sub-weight represents the inverse of this sub-sampling ratio in clusters where sub-sampling has occurred.

Stabilization Weight
Sample stabilization is also used to address problems with sample size growth. Cluster sub-sampling addressed isolated growth in relatively small areas whereas sample stabilization accommodates the slow sample growth over time that is the result of a fixed sampling rate along with a general increase in the size of the population. Sample stabilization is the random dropping of dwellings from the sample in order to maintain the sample size at its desired level. The basic weight is adjusted by the ratio of the sample size, based on the fixed sampling rate, to the desired sample size. This adjustment factor is known as the stabilization weight. The adjustment is done within stabilization areas defined as dwellings belonging to the same employment insurance economic region and the same rotation group.

Non-response
For certain types of non-response (i.e. household temporarily absent, refusal), data from a previous month's interview with the household if any, is brought forward and used as the current month's data for the household.

In other cases, non-response is compensated for by proportionally increasing the weights of responding households. The weight of each responding record is increased by the ratio of the number of households that should have been interviewed, divided by the number that were actually interviewed. This adjustment is done separately for non-response areas, which are defined by employment insurance economic region, type of area, and rotation group. It is based on the assumption that the households that have been interviewed represent the characteristics of those that should have been interviewed within a non-response area.

Labour Force Survey Sub-weight
The product of the previously described weighting factors is called the LFS sub-weight. All members of the same sampled dwelling have the same sub-weight.

Sub-provincial and Province–Age–Sex Adjustments
The sub-weight can be used to derive a valid estimate of any characteristic for which information is collected by the LFS. However, these estimates will be based on a frame that contains some information that may be several years out of date and therefore not representative of the current population. Through the use of more up-to-date auxiliary information about the target population, the sample weights are adjusted to improve both the precision of the estimates and the sample's representation of the current population.

Independent estimates are available monthly for various age and sex groups by province. These are population projections based on the most recent census data, records of births and deaths, and estimates of migration. In the final step, this auxiliary information is used to transform the sub-weight into the final weight. This is done using a calibration method. This method ensures that the final weights it produces sum to the census projections for the auxiliary variables, namely totals for various age-sex groups, economic regions, census metropolitan areas, rotation groups, household and economic family size. Weights are also adjusted so that estimates of the previous month's industry and labour status estimates derived from the present month's sample, sum up to the corresponding estimates from the previous month's sample. This is called composite estimation. The entire adjustment is applied using the generalized regression technique.

This final weight is normally not used in the weighting for a supplement to the LFS. Instead, it is the sub-weight that is used, as explained in the following paragraphs.

11.2 Weighting Procedures for the Canadian Community Health Survey

As described in Section 5.2.3, the CCHS Cycle 3.1 had recourse to three sampling frames for its sample selection: an area frame of dwellings acting as the primary frame and two frames formed of telephone numbers complementing the area frame. Since only minor differences differentiate the two frames formed of telephone numbers in terms of weighting, they are treated together. They are referred to as being part of the telephone frame.

The weighting strategy for the CCHS was developed separately for the area frame and the telephone frame. The weights from these two frames were then combined into one set of weights during the integration stage in the CCHS weighting process. It is worth noting that only individuals from the area frame were used to construct the sample for the SCAL in the territories. Consequently, we present below only the weighting process for this frame.

Weighting of the area frame sample

A0 – Initial weight
Since the mechanism established for the LFS was used to select the area frame sample, the initial weights had to be computed with respect to that mechanism. First, within each stratum defined by the LFS, clusters (primary sampling units) are selected with probabilities proportional to the number of households (based on 2001 Census counts). Next, dwellings are sampled within each selected cluster using systematic random sampling. The product of the probabilities for each of these selections represents the overall probability of selection, and the inverse of that probability is used as the CCHS Cycle 3.1 initial weight.

A1 – Sample increase
Some modifications were made to the default LFS mechanism at the time of sample selection for the CCHS Cycle 3.1. The LFS design provides approximately 60,200 dwellings nationally, but the CCHS Cycle 3.1 requirements in terms of sample size were higher for some regions. Modifications were made in order to obtain extra sample within those health regions (HR) requiring more sample. More specifically, these modifications consisted of repeating the sampling process within all selected clusters of the HR. This had the effect of boosting the sample size and had to be taken into consideration in the weighting by adjusting the probability of selection. An adjustment factor, A1, representing the sample increase rate, was calculated. However, a sample increase was not required in every HR. In some regions, the LFS design provided more sample than needed by the CCHS Cycle 3.1. For those regions, the adjustment factor represents a sample decrease instead of representing a sample increase. The initial weight, A0, is multiplied by this adjustment factor, resulting in weight A1.

A2 – Stabilization
In some HRs, increasing the sample, as described in the previous paragraph, resulted in a significantly larger sample than necessary. Stabilization was therefore instituted to bring the sample size back down to the desired level. The stabilization process consisted of randomly sub-sampling dwellings at the HR level. An adjustment factor representing the effect of this stabilization was calculated in order to adjust the probability of selection appropriately. This factor, multiplied by weight A1, produces weight A2.

A3 – Removal of out-of-scope units
Among all dwellings sampled, a certain proportion is identified during collection as being out-of-scope. Dwellings that are demolished or under construction, vacant, seasonal or secondary, and institutions are examples of out-of-scope cases for the CCHS Cycle 3.1. These dwellings were simply removed from the sample, leaving only a sample that consisted of in-scope dwellings. These dwellings maintain the same weight as in the previous step, which is now called A3.

A4 – Household non-response
During collection, a certain proportion of interviewed households inevitably resulted in non-response. This usually occurs when a household refuses to participate in the survey, provides unusable data, or cannot be reached for an interview. Weights of the non-responding households were distributed using response propensity classes to the responding households. The CHAID (Chi-Square Automatic Interaction Detector) algorithm, available in Knowledge Seeker7, was used to identify the best characteristics to divide the sample into groups that were dissimilar with respect to response/non-response. Note that the groups were formed independently within each HR. Since the information available for non-respondents is limited, only characteristics such as collection period and a rural/urban indicator could be used for the creation of the classes. Analysis concluded that the rural/urban indicator was the most significant characteristic. The rural/urban indicator was also significant (with four periods: January to March, April and May, June to August, and September to November) for the creation of classes in each HR. An adjustment factor was calculated within each class as follows:

$$\frac{sum\,of\;\; weights\; A3\;\; for\; all\; households}{sum\,of\;\; weights\; A3\;\; for\; all\; responding\; households}$$

Weight A3 for responding households was therefore multiplied by this adjustment factor to produce weight A4. For the CCHS, non-responding households are taken out of the weighting process from this point. For the SCAL, the initial weight used was the final weight from the area frame, adjusted for the fact that non-responding households were excluded from the frame, while households where non-response was at the person level were kept.

A5 – Creation of person-level weight
Since persons are the desired sampling units for the CCHS, the household level weights computed to this point need to be converted down to the person level. This weight is obtained by

multiplying weight A4 by the inverse of the probability of selection of the person selected in the household. This gives the weight A5. As mentioned earlier, the probability of selection for an individual changes depending on the number of people in the household and the ages of those individuals. See Section 5.2 for more details.

A6 – Person non-response
A CCHS Cycle 3.1 interview can be seen as a two-part process. First the interviewer gets the complete roster of the people living within the responding household. Second, (s)he interviews the selected person within the household. In some cases, interviewers can only get through the first part, either because they cannot get in touch with the selected person, or because that selected person refuses to be interviewed. Such cases are defined as person non-response and an adjustment factor must be applied to the weights of respondents to account for this non-response. Using the same methodology as was used in the treatment of household non-response, the adjustment was applied within classes based on characteristics available for both respondents and non-respondents. All characteristics collected when creating the roster of household members were in fact available for the creation of the classes. The CHAID algorithm, available in Knowledge Seeker7, was used to define the classes. Note that groups were formed independently within each HR. Depending on the HR, the following characteristics were used to form the adjustment classes: household size groups, urban/rural indicator, collection period, number of persons 12 years of age or older, living arrangement, sex, marital status, and age, resulting in the following adjustment factor:

$$\frac{sum\ of\ weights\ A5\ for\ all\ selected\ persons}{sum\ of\ weights\ A5\ for\ all\ responding\ selected\ persons}$$

Weight A5 for responding persons was multiplied by the above adjustment factor to produce weight A6. Non-responding persons were dropped from the weighting process from this point onward.

Since this adjustment was the last one necessary for the sample drawn from the area frame, weight A6 represents the ***final area frame weight***. This weight is later integrated with the final weight of the telephone frame to create the final CCHS Cycle 3.1 weight.

Since the SCAL sample was selected entirely from the area frame, the initial weight used is the weight obtained from step A6. However, since the sample for the SCAL in the territories was selected from the households who responded to the CCHS and the person selected for the SCAL was not necessarily the same respondent as for the CCHS, the initial weight must be a household weight, and must take into account the non-response of persons in the CCHS. For this reason, the initial weight corresponds to the final weight from the area frame of the CCHS divided by an adjustment for the random selection of an individual from the CCHS household.

11.3 Weighting Procedures for the Survey of Canadian Attitudes toward Learning

The principles behind the calculation of the weights for the SCAL are identical to those for the LFS and the CCHS. However, further adjustments are made to the LFS sub-weights and the household weights of the CCHS in order to derive a final weight for the individual records on the SCAL microdata file.

The different adjustments were the same for individuals selected from the LFS frame and those selected from the CCHS frame, with exception of the 3rd adjustment which is applicable only to the provinces, that is, to the individuals selected from the LFS frame.

1) Probability of selection

The first adjustment consists of adjusting the weight of households selected in the SCAL sample to account for the weight of households that were not selected. This adjustment is made by multiplying within each strata the weights of the selected households by a factor equal to the inverse of their probability of selection, and a weight of 0 is assigned to the other households.

2) Number of persons in the household
The second adjustment aims at accounting for the size of the household selected in order to produce estimates in terms of persons as opposed to households. For this, we multiply the weight obtained in the first adjustment by the number of persons in the target population that are present in the household.

3) Number of rotation groups
This adjustment applies only to the individuals selected in the provinces. Three rotation groups, that is one half of the six needed rotations of the LFS, were used to construct the sample for the SCAL in the provinces. The third adjustment consists of accounting for this selection by multiplying the weights obtained from the previous adjustment by 2.

4) Adjustment for non-response
The fourth adjustment attempts to account for non-response during the survey. For this, we begin by creating groups of individuals (respondents and non-respondents) who had similar probabilities of responding. Then, in each group, we multiplied the weight of individual who responded by a factor equal to:

$$\frac{sum\ of\ the\ weights\ of\ respondents + sum\ of\ the\ weights\ of\ non-respondents}{sum\ of\ weights\ of\ respondents}$$

5) Post-stratification
The last adjustment aims at ensuring that totals produced using data collected in this survey correspond with known demographic estimates. The selected respondents in the provinces were grouped by province, age grouping and gender, while those in the territories were grouped by territory and gender. In each category, we multiplied the weight obtained in the preceding adjustment by a factor equal to:

$$\frac{size\ of\ the\ population\ in\ a\ given\ category}{sum\ of\ the\ weights\ of\ respondents\ in\ this\ category}$$

The resulting weight WTPS is the final weight which appears on the SCAL microdata file.

12.0 Questionnaire

The Survey of Canadian Attitudes toward Learning (SCAL) questionnaire was used in April and May, 2006 to collect the information for the survey. The file SCAL2006_QuestE.pdf contains the English questionnaire.

13.0 Record Layout with Univariate Frequencies

See SCAL2006_Share_CdBk.pdf for the record layout with univariate counts.