Microdata User Guide

Access and Support to Education and Training Survey

2008

Statistics Canada Statistique Canada

Canada

Table of Contents

1.0 Introduction

The Access and Support to Education and Training Survey (ASETS) was conducted by Statistics Canada from June to October 2008, with the cooperation and support of Human Resources and Skills Development Canada (HRSDC). 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

Tourism and the Centre for Education Statistics Division
Client Services
Telephone: 1-800-307-3382; 613-951-7608
Fax: 613-951- 4441
E-mail: educationstats@statcan.gc.ca

2.0 Background

The Access and Support to Education and Training Survey (ASETS) is a new voluntary survey of approximately 72,000 households. The survey's objective is to assess education and training demand in Canada in the context of lifelong learning. The ASETS was funded by HRSDC.

The main survey includes children (age 0 to 17), for whom a "person most knowledgeable" (PMK1), was interviewed, or main respondents (age 18 to 64), from whom information was collected directly. A main respondent is either a youth (age 18 to 24), or an adult (age 25 to 64). Collection for the main survey took place between June 16 and October 6, 2008. A secondary set of interviews started July 11, 2008 and ended October 10, 2008, and collected information from a "person most knowledgeable" (PMK2) of youths. Youths and their responding PMK2s were asked for permission to link their respective survey responses.

The ASETS brings together three previously-conducted surveys that addressed issues relating to antecedents and determinants to access to postsecondary education, the role of student loans and savings in the financing of postsecondary education and participation in adult education and training. The ASETS replaces the Survey of Approaches to Educational Planning (SAEP), the Post-secondary Education Participation Survey (PEPS) and the Adult Education and Training Survey (AETS), which were last conducted in 2002 and 2003. Bringing together the three surveys realizes economies of scale in the management, development, implementation and dissemination of the survey.

The ASETS expands the analytical potential of the three separate surveys, in terms of the content that can be addressed, the sample sizes that can be afforded and by allowing the study of correlates across all three themes within the context of lifelong learning. The data collected by the ASETS will help to monitor preparedness and access to education, evaluate the effectiveness of government education-related programs and develop policies to deal with the training needs of Canadians.

3.0 Objectives

The survey measures and provides indicators of:

- academic hopes and expectations of parents/guardians for their children
- parents/guardians values and attitudes towards education in general, and postsecondary education in particular
- proportion of the population aged 0 to 17 who currently have savings set aside for their postsecondary education by parents/guardians or other family members
- proportion of the population aged 0 to 17 whose parents/guardians plan to start saving at a later date
- saving strategies
- contributions to the Registered Education Savings Plan (RESP)
- expectations for child's contributions to their own postsecondary education including, working while in high school or postsecondary, loans, grants, scholarships, etc.
- characteristics of families that are not and will not be saving for their child's postsecondary education, as well as main reason for not saving
- access to postsecondary education
- the impact of student loan on accessibility of postsecondary education
- the impact of parental socio-economic status on postsecondary accessibility
- the characteristics of postsecondary programs pursued
- the mechanisms through which students finance postsecondary education
- the accessibility of student loans, use of student loans and student indebtedness for those who have initiated postsecondary studies
- the awareness of the Canada Education Savings Program
- the adequacy of student financing through examination of tuition fees, other education costs and major monthly expenses of current students
- the socio-demographic characteristics of those who are: not participating in postsecondary education, have/are participating in postsecondary and have/had a student loan or those who have/are participating in postsecondary and have not had a student loan
- the incidence and intensity of adults' participation in job-related formal training
- the profile of the employer support for job-related formal training
- aspects of job-related training activities such as: training provider, expenses, financial support, motivations, outcomes and difficulties experienced while training
- the barriers preventing individuals from participating in formal training they want or need to take
- the reasons behind adults' lack of participation and interest in formal training

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 used.

Bursary
A monetary award to assist a student in the pursuit of his/her studies based on financial need.

Canada Education Savings Grant (CESG)
A grant offered by the Government of Canada to encourage parents, family and friends to save for a child's education after high-school.

Canada Education Savings Program (CESP)
A grant from the Government of Canada paid directly into a beneficiary's Registered Education Savings Plan (RESP). It adds 20% to 40% in contributions made into an RESP on behalf of an eligible beneficiary each year. Together with the accumulating investment income, the grant will be available to be paid to the student as part of Educational Assistance payments when he/she goes to postsecondary school.

Canada Learning Bond (CLB)
A government of Canada grant to help modest-income families start saving for their child's education after high school.

Canadian citizen
A Canadian citizen is a person who was born in Canada, who was born outside Canada and one of the parents was a Canadian citizen, or who became a Canadian citizen through the naturalization process (e.g. a landed immigrant who has been granted citizenship).

Covered by union contract or collective agreement (not union member)
Many employees work in jobs covered by a union contract, but for one reason or another they have not joined the union, and have not signed a union card due to, for example, personal or religious reasons.

Debt reduction
A debt management measure designed to help manage the repayment of student loans in case of long-term financial difficulties.

Business class immigrant
Immigrants selected for their skills or other assets that will contribute to the Canadian economy (includes skilled workers, investors, entrepreneurs and self-employed persons).

Education tax credit / Tuition fee tax credit / Textbook tax credit
The purpose of these credits is to assist a student by reducing the student's income tax. Any unused portion of the student's education tax credit may be transferred to the student's spouse, a parent or a grandparent, or can be carried forward.

Employees (Note on employees)
Self-employed individuals and their business partners are not considered employees of the company.

Family class
Includes immigrants sponsored by close relatives or family members already living in Canada.

Government sponsored student loan
A student loan is a loan sponsored by the federal government or any provincial/territorial government, which enables the respondent to finance his/her studies. As of March 2001, Canada student Loans come directly from the government of Canada through the National student loan Service Centre. Examples of student loans are: the Canada Student Loan Program (CSLP), the Ontario Student Assistance Program (OSAP), the Quebec Student Loan and Bursaries Program (QSLBP), the Alberta Student Loans Program (ASLP), etc.

Grant
A sum of money given by a government or corporation to a student on the condition that certain terms are accepted or certain engagements fulfilled, as required by the sponsor.

Group scholarship foundation
A company offering pooled investment plans designed to help pay for postsecondary education.

Guaranteed Investment Certificate (GIC)
A certificate showing that a stated amount has been deposited at a financial institution for a specified period of time, usually one to five years, at a stipulated rate of interest. In ordinary circumstances, the deposit may not be withdrawn before maturity.

Independent class immigrant
Immigrants who qualify for certain types of jobs or have other important assets to bring to Canada. They apply on their own or have more distant relatives living in Canada.

Interest Relief
Program designed to provide temporary relief for borrowers who have difficulty repaying their Canada Student Loans. After a grace period, the borrower is not required to make any payments on the loan and the federal government pays the interest. Applicants are normally approved for six-month periods up to a maximum of 30 months.

In-trust account
An in-trust account is an account with a bank, credit union or trust company in which deposits are made as a trust for the benefit of the beneficiary, even if no formal trust agreement exists.

Landed immigrant
A landed immigrant is a person who has been granted the right to live in Canada permanently by immigration authorities, but has not yet obtained Canadian citizenship. These persons are referred to as “permanent residents” under the *Immigration Act.*

Loan remission
The CSLP does not offer loan remissions, but most provinces do offer this debt management measure.

Paid worker
A paid worker is someone who works for others (an employee), i.e., works for an employer, and receives a wage or salary.

Permanent job
A permanent job is one that is expected to last as long as the employee wants it, and as long as business conditions permit. A job that is not permanent is one that has a predetermined date on which it will end or will end as soon as the specified project is completed.

Program
A program is a selection of courses or a combination of courses usually chosen from a syllabus, calendar or a list. Courses within a program are usually taken for credit towards a degree, diploma or certificate. A program is comprised of the courses taken to receive a certificate, diploma and/or degree and must take three months or longer to complete.

Refugees
Refugees are persons seeking protection in Canada.

Registered Education Savings Plan (RESP)
An RESP is a tax-sheltered means of saving to finance a child's postsecondary education. Once the child starts postsecondary studies, the savings in the plan are used to pay tuition or other expenses related to his/her studies.

Registered Retirement Savings Plan (RRSP)
An RRSP is a capital accumulation program designed to encourage saving for retirement. Contributions are tax-deductible within certain limits.

Required fees
These are fees which are paid to the institution and are separate from the course fees. These required fees cover such things as administrative costs, gymnasium costs, maintenance costs, use of laboratories, use of library, membership to the students association and other various expenses paid to the institution.

Saving bonds
Government bonds sold to the general public yielding variable interest.

Scholarship
A monetary award to assist a student in the pursuit of his/her studies usually based on outstanding academic achievement.

Self-employed worker
Someone who owns and operates a business and is paid directly by the client.

Student loan interest tax credit
A claim on the amount of interest paid on the student loan can only be made by the student or a person related to the student who has paid this interest.

Trade/vocational school institute or publicly funded technical institute
The difference between this category and the category "a private training institute or a private business school" is mostly based on the "public" versus "private" nature if the institution.

Trust fund
A fund consisting of assets belonging to a trust held by trustees for beneficiaries who receive money under certain conditions.

Tuition
Tuition is the amount of money charged by an educational institution for instruction.

Union member
To be a union member, the respondent must have actually joined a union, and signed a union card.

Unpaid family worker
Someone who works without pay on a farm or in a business owned and operated by another family member living in the same household.

Volunteer work
Volunteering is defined as doing unpaid activities for groups or organizations such as schools, religious groups, food banks and/or community associations.

5.0 Survey Methodology

This chapter describes the target population, the sampling frame and the sample design of the Access and Support to Education and Training Survey (ASETS).

5.1 Target Population

The target population for the ASETS is comprised of all Canadian residents aged less than 65 years old, excluding individuals residing in the three territories in the North and excluding individuals residing in institutions.

5.2 Sampling Frame

The ASETS frame is a telephone list frame, constructed from two sources of telephone numbers:

- the 2006 Canadian Census of Population telephone numbers and
- a list of ”residential” telephone numbers compiled from telephone company administration files.

The frame was constructed in several steps:

- obtained a file containing all the Census households;
- dropped all Census households with no valid telephone number (7% of households), with duplicate telephone numbers (1.5%), or with imputed birth dates (2%);
- added telephone numbers from the administration files: only telephone numbers which were missing from the Census portion of the frame were added.

The ASETS frame consists of a list of distinct telephone numbers with no duplicates. It can therefore be treated as a single frame (as opposed to a dual frame) even though it was constructed from two sources.

5.2.1 Population Coverage

The ASETS survey population does not include individuals without telephones or with cellular phones only. As well, it does not include individuals whose current telephone number belonged to households whose members were all aged 65 and over according to the Census. Lastly, the frame does not include households whose telephone number is missing from both the 2006 Census telephone list and from the administration files.

Frame coverage is discussed further in Section 8.2.1.

5.3 Sample Design

5.3.1 Two-phase design of the Census Portion

The 2006 Census was conducted using a short questionnaire (Form 2A) and a long questionnaire (Form 2B/2D). In most regions of the country, one in five households were selected at random to receive the 2B/2D questionnaire; the other households received the 2A questionnaire. The Census portion of the ASETS sample is a sub-sample of the Census 2B/2D sample. Therefore this portion of the sample has a two phase design: the first phase is the Census 2B/2D sample, and the second phase is the ASETS sub-sample.

Selecting a sub-sample from the Census households which completed the 2B/2D long form questionnaire had the advantage that more Census variables were available to perform the non-response weighting adjustments.

5.3.2 Stratification

The telephone numbers on the ASETS frame were stratified by province, and each province was further sub-stratified into six strata, as defined in the following table:

Telephone Source	Stratum Name	Description
Census	Youth only	Households composed of youths aged 18 to 24 only
	Youth mixed	Households composed of youths and other age groups
	Children mixed	Households composed of children aged 0 to 17 and no youth
	Adults only	Households composed of adults aged 25 to 64 only
	Seniors only	Households composed of seniors aged 65 and over only
Administrative	Administrative	Telephone numbers from administration files

The first five strata (Youth only, Youth mixed, Children mixed, Adults only and Seniors only) are strata for the Census portion of the frame. The Census birth date was used to calculate the age of each household member and to classify the households into one of the five strata.

The Administrative stratum is a separate stratum for the telephone numbers obtained from administration files.

5.3.3 Telephone Number Selection

For the Census portion of the frame, the sample of telephone numbers were selected using systematic probability proportional to size (PPS) sampling within each stratum from the Census 2B sample. The size measure used was the Census weight for the household. The PPS sampling was used so that all households within the same stratum would end up with equal design weights.

For the Administrative strata, the telephone numbers were selected using simple random sampling without replacement (SRSWOR).

The telephone numbers of the selected sample were sent to the field for collection. The interview was conducted with the household reached by the selected telephone number, regardless of whether it was the same household as on the frame or still had the same age composition.

5.3.4 Person Selection

Upon contact with a household, a roster listing all household members and their age was collected, and a household member aged between 0 and 64 was randomly selected for the survey. In order to meet the survey requirements, the probability of selection

depended on age: youths aged 18 to 24 were eight times more likely of being selected than adults aged 25 to 64, and five times more likely than children aged 0 to 17. If all members of the household were aged 65 and over, no household member was selected and the interview ended.

5.4 Sample Size

The ASETS sample consisted of approximately 72,000 telephone numbers. The following table shows how the sample was allocated to the household type strata.

Stratum	Sample Size	Sampling Fraction
Youth Only	1,360	0.0108
Youth Mixed	19,640	0.0108
Children Mixed	17,000	0.0072
Adults Only	14,000	0.0030
Seniors Only	0	0.0000
Admin	20,000	0.0070
Total	**72,000**	

As shown in the above table, households containing youths according to the Census data were over-sampled, whereas those composed only of adults were under-sampled. Households composed only of seniors were not sampled. The sample allocation was driven by the sample size requirements for the different age groups.

Kish allocation was used to allocate the sample between the provinces. A Kish parameter of 0.67 was used, which gives an allocation which is a compromise between equal allocation and proportional allocation. The following table shows the sample size by province.

Province	Sample Size
Newfoundland and Labrador	3,811
Prince Edward Island	3,725
Nova Scotia	4,004
New Brunswick	3,902
Quebec	12,958
Ontario	20,864
Manitoba	4,140
Saskatchewan	4,022
Alberta	6,879
British Columbia	7,695
Canada	**72,000**

6.0 Data Collection

Collection for the Access and Support to Education and Training Survey (ASETS) took place between June 16 and October 10, 2008 and was done by computer-assisted telephone interview (CATI). An introductory letter was mailed to the selected households approximately two weeks before data collection began.

The CATI system has a number of generic modules, which can be quickly adapted to most types of surveys. A front-end module contains a set of standard response codes for dealing with all possible call outcomes, as well as the associated scripts to be read by the interviewers.

A standard approach set up for introducing the agency, the name and purpose of the survey, the survey sponsors, how the survey results will be used, and the duration of the interview was used. We explained to respondents how they were selected for the survey, that their participation in the survey is voluntary, and that their information will remain strictly confidential. Help screens were provided to the interviewers to assist them in answering questions that are commonly asked by respondents.

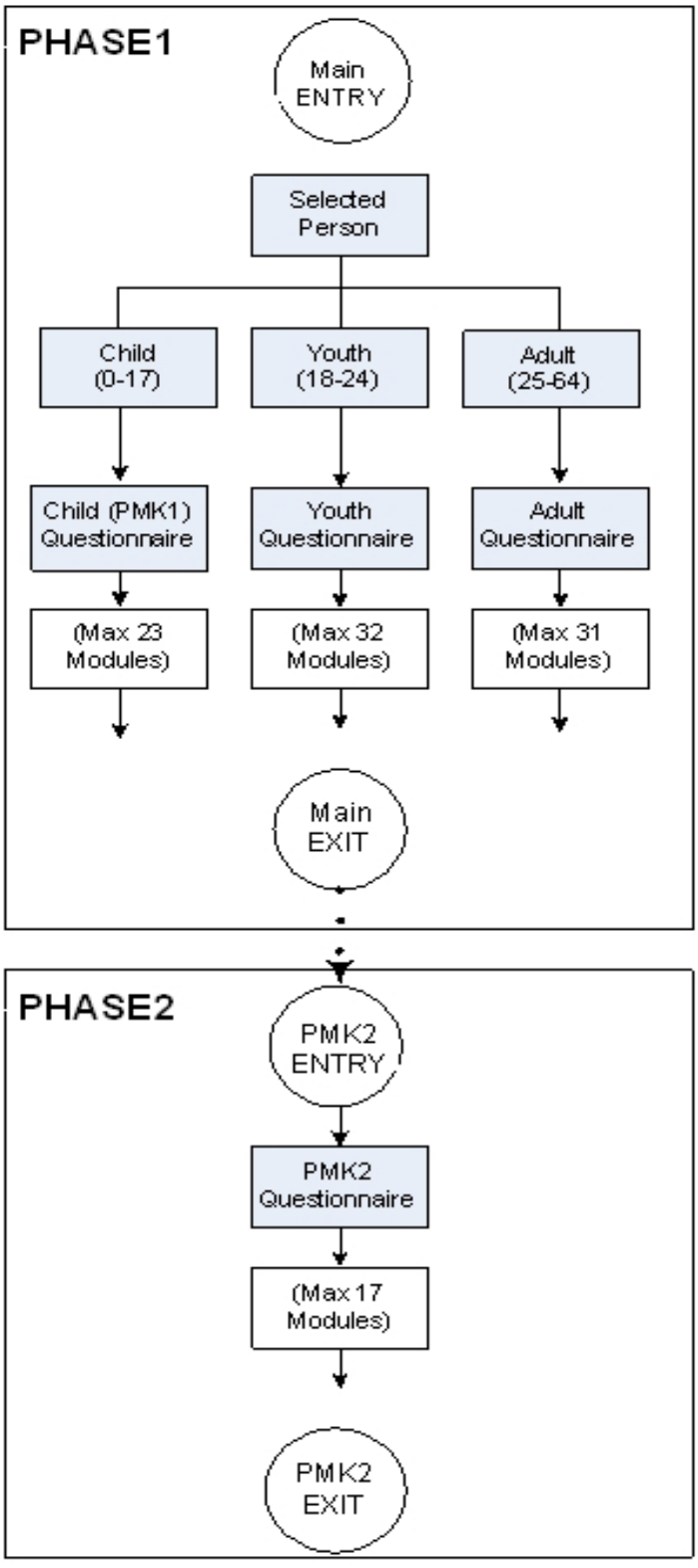
Collection for the ASETS was carried out in two phases. Phase 1 collected information about each selected child (age 0 to 17), for whom a "person most knowledgeable" (PMK1) was interviewed, or a main respondent (age 18 to 64), from whom information was collected directly. Phase 2 collected information from a "person most knowledgeable" (PMK2) for main respondents age 18 to 24. Each collection phase used a different CATI application.

As part of Phase 1, in each selected dwelling, standard socio-demographic information about all household members was first obtained from an adult member of the household. Such "proxy" reporting is used to avoid the high cost and extended time requirements that would be involved in the repeat calls necessary to obtain information directly from each respondent.

Upon completion of the demography module, a random selection was carried out for Phase 1 (ASETS Main), and the interviewer asked to speak to the selected person, or the PMK1 (in case a child had been selected). If this person was not available, the interviewer arranged for a convenient time to phone back.

Once the ASETS interview for a youth was completed, we asked the youth to provide us with contact information for a PMK2, i.e., a person most knowledgeable about the youth's financial preparedness for postsecondary education. Phase 2 of the collection consisted of administering the ASETS questionnaire to PMK2s, once telephone contact had been established.

The requirement for a second phase presented additional challenges to the programming of the CATI application and its implementation during data collection. Some information obtained during Phase 1 had to be made available for Phase 2 and was transferred through the sample file. This process was not always successful (see Section 8.2.3 on Data Processing).



The data collection was conducted by specialized staff at Statistics Canada offices in Edmonton, Sturgeon Falls, Toronto, Halifax, Winnipeg and Sherbrooke. The workload and interviewing staff within each office was managed by a project manager. The automated scheduler used by the CATI system ensured that cases were assigned randomly to interviewers and that cases were called at different times

of the day and different days of the week to maximize the probability of contact. There were a maximum of 25 call attempts per case; once the maximum was reached, the case was reviewed by a senior interviewer who determined if additional calls would be made.

6.1 Supervision and Quality Control

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

6.2 Non-response

Interviewers were instructed to make all reasonable attempts to obtain ASETS interviews with members of eligible households. For individuals who at first refused to participate in the ASETS, a letter was sent from the Regional Office to the dwelling address stressing the importance of the survey and the household's cooperation. For cases in which the timing of the interviewer's call was inconvenient, an appointment was arranged to call back at a more convenient time. For cases in which there is no one home, several call backs were made. Under no circumstances were sampled dwellings replaced by other dwellings for reasons of non-response.

To increase the response rate, the collection period was extended by two weeks beyond what had originally been planned. An appealing postcard was created and mailed from head office to non-respondent households stressing the importance of the survey and inviting selected households to participate. Although this attempt at obtaining interviews appeared to be successful, a number of non-respondent households still remain.

7.0 Data Processing

The main output of the Access and Support to Education and Training Survey (ASETS) is a "clean" microdata file. This chapter presents a brief summary of the processing steps involved in producing this file.

The output of the ASETS has been stored in two separate data files: a **main** file and a **course** file.

The table below lists all the ASETS modules and the population covered by each module. A main respondent is either a youth (age 18 to 24) or an adult (age 25 to 64) and can be identified on the main file by setting the derived variable RESMR to 1. Records with child/PMK1 data can be retrieved by setting RESPMK to 2, whereas RESPMK = 3 identifies records with data from youths and their PMK2. The name of variables (including imputation flags) for which information pertains to or is obtained from a PMK has the letter *P* at the end. For instance, DWELCUR contains information on the type of dwelling for the current residence of the main respondent, whereas DWELCURP, which has the same meaning as DWELCUR, contains PMK1 or PMK2 data.

ASETS Blocks		Population		
Abbreviation	**Block Title**	**Main respondent**	**PMK1**	**PMK2**
LP	Information on learning	X		
AL	Attitude towards learning	X	X	X
EM	Most recent employment	X	X	X
SP	Volunteering	X		
EH	High school experience	X		
EA	Postsecondary experience	X		
HE	Highest education	X		
EC	Formal education	X	EC_Q04	EC_Q04
MR	Most recent program	X		
EE	Duration of formal education	X		
EO	Details of formal education	X		
SE/FM	Employer support/Employment situation	X		
CE	Costs of formal education	X		
ES	Financial support – savings	X		
ER	Financial education repayable loans	X		
EN	Financial support - non-repayable	X		
EJ	Current student with loan	X		
EK	Current student without loan	X		
CR1/PN	Non-formal education/Course name	X		
CR2	Job-related courses	X		
CN	Job-related courses (on Course file)	X		
CO/FM	Cost non-formal/Employment situation (on Course file)	X		
BE	Barriers	X		
FB	Family background	X	X	
CC	Child's conditions		X	
SA	Child school attendance		X	

ASETS Blocks		Population		
Abbreviation	**Block Title**	**Main respondent**	**PMK1**	**PMK2**
CS	Child school experience		X	
PH	Parental expectations		X	X
PP	Parental involvement		X	
CT	Computer use	X	X	
IM	Immigration	X	X	
LG	Language	X	X	X
LD	Disabilities	X		
FP	Financial planning		X	X
FO	Financial preparedness if in PSE		X	X
SO	Parents saving for other children		X	X
SS	Parents savings		X	X
SW	Savings withdrawals		X	X
NS	No savings		X	X
PA	Savings awareness		X	X
HL	Principal residence	X	X	X
EL	Student loan impact	X		
IE	Household expenditures	X	X	X
ET	Financial support	X	X	X
IN	Income	X	X	X

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 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

Information collected from the respondents often requires coding so that the data can be better analyzed. In the case of the ASETS, several coding systems were used.

Classification of instructional programs (CIP)
Coding information relating to instructional programs was performed using the CIP codes (see Appendix A).

Industry and Occupation
The North American Industry Classification System (NAICS) 2007 was used to code the employer's type of business (see Appendix B). The National Occupational Classification for Statistics (NOC-S) 2006 was used to code information relating to the kind of work and main activity reported by respondents (see Appendix C).

Country codes
The 2006 Census country code set was used to code the country of origin (see Appendix D).

Language codes
The 2006 Census language code set was used to code the language spoken most often in the household (see Appendix E).

Ethnic origin
The 2006 Census ethnic origin code set was used to code the ethnic and cultural origins (see Appendix F).

Other - Specify
Questions which contain a list of answer categories often contain "Other - Specify" as the final category. The text from these questions is captured. These write-ins are examined and may be recoded into one of the existing categories. If the write-in is reflected in one of the existing categories to the question, the appropriate category is set to "*Yes"* and the *"Other - Specify"* is set to "*No"*. Also, new categories may be added if there is a large number of write-ins which can be categorized together.

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.

In the case of the ASETS, imputation was used to fill in missing data for item non-response. Donor imputation was used to impute missing PMK2 data. Further information on the imputation process is given in Chapter 8.0.

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. The following is a list of derived variables that have been created. Please see the codebook for the complete list of derived variables and their values.

IDENTIFICATION VARIABLES	
Variable	**Description**
RESMR	Type of respondent (youth 18 to 24, adult 25 to 64)
RESPMK	Type of respondent (PMK of child 0 to 17 or PMK of youth 18 to 24)
PMKHHLD	Location of the PMK
MRSTERRP	Identification of error in PMK2 marital status coding as used during collection

SOCIO-DEMOGRAPHIC VARIABLES	
Variable	**Description**
DPROV	Province code for respondent's current address
REGION	Region derived from province of current residence of the selected respondent
DWELCUR	Type of dwelling for current residence of main respondent
DWELCURP	Type of dwelling for current residence of PMK1 or PMK2
DWELSTU	Type of dwelling for residence during school year
DVAGE	Age of selected respondent
AGEDEC	Age of selected respondent as of December 31, 2007
IMMAGE	Age of selected respondent at time of immigration
LANGUAGE	Language spoken most often in the household
RELSELP	Relationship of PMK1 or PMK2 to the selected respondent
RELSELSP	Relationship of PMK1 spouse or partner to respondent
CHDLT25	Number of PMK's children 24 years of age or younger who live with the PMK in the selected household
FAMSCHD	Family structure of the selected child
FAMSTPA	Family structure – past
MOM_ED	Highest level of education of mother living in the selected household
DAD_ED	Highest level of education of father living in the selected household
PAR_ED	Highest level of education of parents living in the selected household

SOCIO-DEMOGRAPHIC VARIABLES	
Variable	**Description**
PMK1EDP	Highest level of education of PMK1
OTHERED	Highest level of education of other parent or guardian living outside of the selected household
RESPINC	Income of main respondent – grouped
RESPINCP	Income of PMK – grouped
TOTINC	Main respondent and their spouse or partner's combined income - grouped
TOTINCP	PMK and their spouse or partner's combined income - grouped

OCCUPATION AND INDUSTRY CODING (MAIN FILE)	
Variable	**Description**
EM1CLSW	Class of worker, main respondent
EM1CLSWP	Class of worker, PMK
EM2CLSWP	Class of worker, spouse or partner of PMK
EM1COCC	Unit group for National Occupation Classification for Statistics, main respondent (4-digit level)
EM1COCCP	Unit group for National Occupation Classification for Statistics, PMK (4-digit level)
EM2COCCP	Unit group for National Occupation Classification for Statistics, spouse or partner of PMK (4-digit level)
EM1CIND	Industry group (4-digit level), main respondent
EM1CINDP	Industry group (4-digit level), PMK
EM2CINDP	Industry group (4-digit level), spouse or partner of PMK
EM1NA18	Industry grouping, main respondent
EM1NA18P	Industry grouping, PMK
EM2NA18P	Industry grouping, spouse or partner of PMK
EM1NO2	Major group for occupation (grouped to first 2 digits of NOC-S code), main respondent
EM1NO2P	Major group for occupation (grouped to first 2 digits of NOC-S code), PMK
EM2NO2P	Major group for occupation (grouped to first 2 digits of NOC-S code), spouse or partner of PMK
EM1NA2	Sectors (grouped to first 2 digits of NAICS code), main respondent
EM1NA2P	Sectors (grouped to first 2 digits of NAICS code), PMK
EM2NA2P	Sectors (grouped to first 2 digits of NAICS code), spouse or partner of PMK
EM1NA3	Sub-sectors (grouped to first 3 digits of NAICS code), main respondent
EM1NA3P	Sub-sectors (grouped to first 3 digits of NAICS code), PMK
EM2NA3P	Sub-sectors (grouped to first 3 digits of NAICS code), spouse or partner of PMK

OCCUPATION AND INDUSTRY CODING (MAIN FILE)	
Variable	**Description**
EM1SECT	Public or private sector, main respondent
EM1SECTP	Public or private sector, PMK
EM2SECTP	Public or private sector, spouse or partner of PMK
EM1NO3	Minor group for occupations (grouped to first 3 digits of NOC-S code), main respondent
EM1NO3P	Minor group for occupations (grouped to first 3 digits of NOC-S code), PMK
EM2NO3P	Minor group for occupations (grouped to first 3 digits of NOC-S code), spouse or partner of PMK
EM1NO10	Broad occupational category first digit of the NOC-S code, main respondent
EM1NO10P	Broad occupational category, first digit of the NOC-S code, PMK
EM2NO10P	Broad occupational category, first digit of the NOC-S code, spouse or partner of PMK
FM1CIND	Industry group (4-digit level) - supported by an employer for the most recent program
FM1COCC	Unit group for National Occupation Classification for Statistics (4-digit level) - supported by an employer for the most recent program
FM1SECT	Public or private sector - supported by an employer for the most recent program
FM1NA2	Sectors (grouped to first 2 digits of NAICS code) - supported by an employer for the most recent program
FM1NA3	Sub-sectors (grouped to first 3 digits of NAICS code) - supported by an employer for the most recent program
FM1NA18	Industry grouping - supported by an employer for the most recent program
FM1NO2	Minor group for occupations (grouped to first 2 digits of NOC-S code) - supported by an employer for the most recent program
FM1NO3	Minor group for occupations (grouped to first 3 digits of NOC-S code) - supported by an employer for the most recent program
FM1NO10	Industry group, first digit of the NOC-S code - supported by an employer for the most recent program

OCCUPATION AND INDUSTRY CODING (COURSE FILE)	
Variable	**Description**
FM2CIND	Industry group (4-digit level) - supported by an employer for course, seminar or workshop
FM2COCC	Unit group for National Occupation Classification for Statistics (4-digit level) - supported by an employer for course, seminar or workshop
FM2SECT	Most recent job held by respondent belongs to the public or private sector - supported by an employer for course, seminar or workshop
FM2NA2	Sectors (grouped to first 2 digits of NAICS code) - supported by an employer for course, seminar or workshop
FM2NA3	Sub-sectors (grouped to first 3 digits of NAICS code) - supported by an employer for course, seminar or workshop
FM2NA18	Industry grouping - supported by an employer for course, seminar or workshop

OCCUPATION AND INDUSTRY CODING (COURSE FILE)	
Variable	**Description**
FM2NO2	Major group for occupations (grouped to first 2 digits of NOC-S code) - supported by an employer for course, seminar or workshop
FM2NO3	Minor group for occupations (grouped to first 3 digits of NOC-S code) - supported by an employer for course, seminar or workshop
FM2NO10	Broad occupational category, first digit of the NOC-S code - supported by an employer for course, seminar or workshop

EDUCATIONAL PLANNING	
Variable	**Description**
PAREXP	Flag to indicate if parent expects child to attend postsecondary education
RESPELCP	Total value of RESP savings by PMK for the selected child at the end of 2007
OTHSAV1P	Total value of other savings by PMK for the selected child at the end of 2007
SELSAVOP	Indicator of who is currently saving for the selected child's postsecondary education
SELWHPYP	Indicator for the selected child whose PMK will pay or help pay for their postsecondary education
SAVESTAP	Household saving status for the selected child's postsecondary education
SAVESTOP	Household saving status for the postsecondary education of children other than the selected child
SAVECHDP	Household saving status by children's location
RESPSTAP	Household RESP savings status for children living in the household
SAVSELCP	Total household savings for selected child at the end of 2007

PARTICIPATION IN EDUCATION	
Variable	**Description**
SECSTAT	Secondary enrolment status as of June 2008
PSSTAT	Postsecondary enrolment status
HEDATT	Highest education attained
AGELFSEC	Age when completed high school or left elementary/high school
AGECOMPS	Age at time of starting first postsecondary program
DATLFSEC	Date left high school (or elementary)
DATHLEV	Date respondent completed highest level of education
MOSECPS	Number of months between completing high school or leaving elementary/high school and starting postsecondary education
MOLFSEC	Number of months between completing high school or leaving elementary/high school and time of the survey (June 30, 2008)
DATSTRPS	Date started first postsecondary program
TYPEMRPG	Type of most recent or current program

PARTICIPATION IN EDUCATION	
Variable	**Description**
GOVSTAT	Government student loan status
TOTEDEX	Total educational expenses for most recent program and during the reference period July 2007 to June 2008
TOTREPS	Total repayable sources of funding for reference period of July 2007 to June 2008
TOTNREPS	Total non-repayable sources of funding for reference period of July 2007 to June 2008

EDUCATION AND TRAINING	
Variable	**Description**
PAIDHRS	Number of paid hours worked per week at their most recent job - main respondent
PAIDHRSP	Number of paid hours worked per week at their most recent job - PMK1 or PMK2
TKPM	Respondent took a job-related program
TKNBPM	Number of job-related programs
TKTA	Respondent took job-related training activities
TKNBTA	Number of job-related courses, workshops or seminars
TAHRPM0	Duration of most recent program, in hours
TAHRCR0	Duration of selected job-related training activity, in hours (on Course File)
TAHRCR1	Duration of first job-related training activity, in hours
TAHRCR2	Duration of second job-related training activity, in hours
TAHRCR3	Duration of third job-related training activity, in hours
TAHRCR4	Duration of fourth job-related training activity, in hours
TAHRCR5	Duration of fifth job-related training activity, in hours
TKHRCR	Duration of job-related training activity - total hours
TEPM	Employer sponsored for most recent program

CLASSIFICATION OF INSTRUCTIONAL PROGRAMS (CIP) - PROGRAMS	
Variable	**Description**
MR1_C01	Classification of Instructional Programs Code, most recent program
MR2_C01	Classification of Instructional Programs Code, second most recent program
MR3_C01	Classification of Instructional Programs Code, third most recent program
MR4_C01	Classification of Instructional Programs Code, fourth most recent program
MR5_C01	Classification of Instructional Programs Code, fifth most recent program
MR6_C01	Classification of Instructional Programs Code, sixth most recent program
MR7_C01	Classification of Instructional Programs Code, seventh most recent program
MR8_C01	Classification of Instructional Programs Code, eighth most recent program

CLASSIFICATION OF INSTRUCTIONAL PROGRAMS (CIP) - PROGRAMS	
Variable	**Description**
MR9_C01	Classification of Instructional Programs Code, ninth most recent program
MR10_C01	Classification of Instructional Programs Code, tenth most recent program
MR1_G01	Grouped Classification of Instructional Programs Code, most recent program
MR2_G01	Grouped Classification of Instructional Programs Code, second most recent program
MR3_G01	Grouped Classification of Instructional Programs Code, third most recent program
MR4_G01	Grouped Classification of Instructional Programs Code, fourth most recent program
MR5_G01	Grouped Classification of Instructional Programs Code, fifth most recent program
MR6_G01	Grouped Classification of Instructional Programs Code, sixth most recent program
MR7_G01	Grouped Classification of Instructional Programs Code, seventh most recent program
MR8_G01	Grouped Classification of Instructional Programs Code, eighth most recent program
MR9_G01	Grouped Classification of Instructional Programs Code, ninth most recent program
MR10_G01	Grouped Classification of Instructional Programs Code, tenth most recent program

CLASSIFICATION OF INSTRUCTIONAL PROGRAMMES (CIP) - COURSES	
Variable	**Description**
CR1_C01	Classification of Instructional Programs Code, most recent course
CR1_C02	Classification of Instructional Programs Code, second most recent course
CR1_C03	Classification of Instructional Programs Code, third most recent course
CR1_C04	Classification of Instructional Programs Code, fourth most recent course
CR1_C05	Classification of Instructional Programs Code, fifth most recent course
CR1_C06	Classification of Instructional Programs Code, sixth most recent course
CR1_C07	Classification of Instructional Programs Code, seventh most recent course
CR1_C08	Classification of Instructional Programs Code, eighth most recent course
CR1_C09	Classification of Instructional Programs Code, ninth most recent course
CR1_C10	Classification of Instructional Programs Code, tenth most recent course
CR1_G01	Grouped Classification of Instructional Programs Code, most recent course
CR1_G02	Grouped Classification of Instructional Programs Code, second most recent course
CR1_G03	Grouped Classification of Instructional Programs Code, third most recent course
CR1_G04	Grouped Classification of Instructional Programs Code, fourth most recent course
CR1_G05	Grouped Classification of Instructional Programs Code, fifth most recent course
CR1_G06	Grouped Classification of Instructional Programs Code, sixth most recent course
CR1_G07	Grouped Classification of Instructional Programs Code, seventh most recent course
CR1_G08	Grouped Classification of Instructional Programs Code, eighth most recent course

CLASSIFICATION OF INSTRUCTIONAL PROGRAMMES (CIP) - COURSES	
Variable	**Description**
CR1_G09	Grouped Classification of Instructional Programs Code, ninth recent course
CR1_G10	Grouped Classification of Instructional Programs Code, tenth most recent course

7.6 Weighting

The principle behind estimation in a probability sample such as the ASETS 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 enrolled in full-time programs at a university 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 and Hit Rates

The following table summarizes the response rates and hit rates to the Access and Support to Education and Training Survey (ASETS).

Province	Telephone numbers sent to collection	Estimated number of in-scope	Hit Rate (%)	Number of respondents	Response Rate (%)	Number of respondents who agreed to share	Share Rate (%)
Newfoundland and Labrador	3,760	2,767	74	1,667	60	1,529	92
Prince Edward Island	3,554	2,465	69	1,567	64	1,468	94
Nova Scotia	3,998	2,834	71	1,934	68	1,814	94
New Brunswick	3,875	2,742	71	1,772	65	1,652	93
Quebec	12,929	8,775	68	5,902	67	5,530	94
Ontario	20,801	14,313	69	9,061	63	8,636	95
Manitoba	4,120	2,757	67	1,867	68	1,768	95
Saskatchewan	3,987	2,895	73	1,806	62	1,727	96
Alberta	6,878	4,674	68	2,875	62	2,749	96
British Columbia	7,656	5,094	67	3,058	60	2,882	94
Canada	**71,558**	**49,305**	**69**	**31,509**	**64**	**29,755**	**94**

The columns in the table are defined as follows:

Telephone numbers sent to collection
These counts are lower than the sample size counts given in Section 5.4. The original plan was to send 72,000 telephone numbers to collection, but 442 were dropped because of overlap with other surveys.

Estimated number of in-scope
When the telephone numbers were returned from collection, the cases were classified as in-scope, out-of-scope or unresolved. The unresolved cases are cases with insufficient data to be classified. The estimated number of in-scope telephone numbers / households is the number of cases that were resolved as in-scope during collection, plus a portion of the unresolved cases.

Hit Rate

$$Hit\ Rate = \frac{Estimated\ \ number\ \ of\ \ in-scope}{Telephone\ \ numbers\ \ sent\ \ to\ \ collection}$$

Number of respondents
The number of cases considered as respondents on the Statistics Canada Master file.

Response Rate

$$\text{Response } Rate = \frac{Number\ \ of\ \ respondents}{Estimated\ \ number\ \ of\ \ in-scope}$$

Number of respondents who agreed to share
The share file contains data for all respondents who gave permission to share their data with Human Resources and Skills Development Canada (HRSDC). This is the number of cases on the share file for HRSDC.

Share Rate

$$\text{Share } Rate = \frac{\text{Respondents who agreed to share}}{\text{Number of respondents}}$$

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

The coverage of the ASETS frame was evaluated by dividing the sum of the ASETS pre-calibration weights by demographic counts. The following table show the results obtained by age group.

Age Group	Coverage (%)
0 to 17	80
18 to 21	79
22 to 24	57
25 to 34	61
35 to 44	77
45 to 54	85
55 to 64	89
Total	**78**

Coverage is lower for the 22 to 34 age group, which can be explained by the higher mobility rates and higher proportion of persons in this age group living in households with cellular phones only.

It was suspected that the coverage of youths living away from their parents was likely to be lower than those living with their parents. This was verified by calculating the proportion of youths living with someone at least 18 years older than themselves using the ASETS respondents, and comparing the same proportion with the Census. The following results were obtained:

Age Group	ASETS (%)	Census (%)
18 to 21	92	84
22 to 24	72	57

The table shows that the ASETS over-represents youths living with a ”parent” (someone at least 18 years older than themselves). The table is based on pre-calibration weights.

Calibration was used to inflate the weights for each age group to demographic totals. It was also used to correct, to some extent, the over-representation of youths living with a parent.

8.2.2 Data Collection

The collection supervisors were trained at head office in Ottawa and personnel from the head office were available to answer questions during the interviewer training sessions. A file containing all questions asked during the training sessions and the answers was produced and distributed to the regional offices

Supervisor and interviewer training consisted of reading the Supervisor’s Manual and Interviewer’s Manual to become familiar with the concepts and definitions of the survey. A description of the background and objectives was provided, as well as a glossary of terms and a set of questions and answers commonly asked by respondents.

8.2.3 Data Processing

Collection for the ASETS was done using a computer-assisted telephone interviewing (CATI) application; therefore, the quality of the data is high. The two major benefits of using a CATI application are:

- ensuring that the correct flow path of the questions is followed and
- the automatic verification of inconsistent responses by using edits within the application.

During collection, the marital status of the PMK2 (person most knowledgeable), who lived in the selected household, was collected through the demographic module in Phase 1 and had to be transferred to Phase 2 of the application. This transfer failed and during the interview the dynamic text concerning the spouse or partner of the respondent did not work properly. The affected questions, of which the most important is a question on a combined income, can be identified using the derived variable MRSTERRP - Flag identifying whether the proper marital status code was used for PMK2 respondents during collection.

Data processing of the ASETS 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 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.

All imputations involved donors that were selected using a score function. For each item non-response or partial non-response records (also called recipient records), we compared certain characteristics to characteristics from all the donors. When the characteristics were the same between a donor and the recipient, a value was added to the score of that donor. The donor with the highest score was deemed the "closest" donor and was chosen to fill in missing pieces of information of the non-respondents. If there was more than one donor with the highest score, a random selection occurred. The pool of donors was made up in such a way that the imputed value assigned to the recipient, in conjunction with other non-imputed items from the recipient would still pass the edits. Note that the lower the imputation rate, the greater the likelihood of finding a donor with characteristics similar to the recipient's.

The following table provides a list of the variables that were imputed and imputation rates for each variable. The column "Records Imputed (%)" is the (unweighted) proportion of records with an imputed value - only records with a value for the variable were included in the calculation (i.e. valid skips were excluded). The column "Estimate Imputed (%)" is the proportion of the total (weighted) estimate that is based on imputed values.

Age of Selected Respondent	Variable Name	Description	Records Imputed (%)	Estimate Imputed (%)
Child 0 to 17	SS_Q09P	Value of Registered Education Savings Plan (RESP) at the end of 2007	33	28
	SS_Q13P	Total contributions to RESP in 2007	26	23
	SS_Q18P	Value of other savings at the end of 2007 (exclude RESPs)	33	29
	SS_Q20P	Total contributions to other savings in 2007	32	24
	IN_Q02P	Person most knowledgeable's (PMK) total income before taxes and deductions from all sources in 2007	26	27
	IN_Q04P	PMK and spouse/partner's total income before taxes and deductions from all sources in 2007	39	41

Age of Selected Respondent	Variable Name	Description	Records Imputed (%)	Estimate Imputed (%)
Youth 18 to 24	EE_Q14	Duration of most recent program	12	12
	CR2Q07_1 to CR2Q07_9 and CR2Q7_10	Duration of job-related courses	7	10
	SS_Q09P	Value of RESP at the end of 2007	45	48
	SS_Q13P	Total contributions to RESP in 2007	40	50
	SS_Q18P	Value of other savings at the end of 2007 (exclude RESPs)	51	58
	SS_Q20P	Total contributions to other savings in 2007	46	53
	IN_Q02P	PMK's total income before taxes and deductions from all sources in 2007	44	51
Adult 25 to 64	EE_Q14	Duration of most recent program	8	7
	CR2Q07_1 to CR2Q07_9 and CR2Q7_10	Duration of job-related courses	8	11

Additional variables were imputed so that the flows to the variables listed in the above table would be correct.

For cases where the selected respondent was aged 18 to 24 (youth), follow-up interviews were conducted with the youth's parent or guardian (person most knowledgeable (PMK2)). There are 1,881 cases where an interview was conducted with the youth, but no data was collected from the PMK2 about the youth. Non-response to the follow-up PMK 2 interviews were handled by donor imputation: youth cases with no PMK2 data were imputed by cases with reported PMK2 data. Cases where all the variables from the PMK2 interview were imputed are identified through an imputation flag called IPMK2P. The following table provides imputation rates for the youth PMK2 data:

Description	PMK2 Data Imputed (%)
Youths living in same household as PMK2	20
Youths living in different household than PMK2	55

Note that the PMK2 data imputation rate is particularly high for youths living in different households than their PMK2. The high rate implies that we were less likely to find donors with characteristics similar to the recipients for these youths.

Additional imputations were performed for two problems with the CATI application. The first problem was that the variables IM_Q15 and IM_Q15P, used to derive the visibility minority flag called VISMIN, were not asked of persons born outside of Canada, the US and Greenland. The visible minority flag was imputed for all these cases using country of birth (IM_C05, IM_C05P) and information from the 2006 Census. The second problem that occurred during collection was that combined income, IN_Q04P, was not asked of the PMK of youths in all cases where the PMK had a spouse/partner, and the PMK and the youth lived in the same household. The imputation for IN_Q04P was performed using 2006 Census data for all cases were the PMK was a parent of the youth, and the PMK and their spouse were opposite-sex partners. An income category, adjusted for reference

year, was imputed using a Census donor with similar characteristics as the ASETS recipient.

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 62.1% of children aged 0 to 17 have PMK1s who hope they will pursue postsecondary studies and who have set aside (possibly with their spouse) savings for the child's postsecondary education. Suppose that this estimate is found to have a standard error of 0.004. Then the coefficient of variation of the estimate is calculated as:

$$\left(\frac{0.004}{0.621} \right) X \ 100 \ \% \ = \ 0.64 \ \%$$

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

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

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

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

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

9.2 Sample Weighting Guidelines for Tabulation

The sample design used for the Access and Support to Education and Training Survey (ASETS) 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 ASETS 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 ASETS.

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. Examples of such estimates would be the number of children whose person most knowledgeable (PMK1) have set aside savings for postsecondary education, or the proportion of adults who took training through distance education. An estimate of the number of persons possessing a certain characteristic may also be referred to as an estimate of an aggregate.

Examples of Categorical Questions:

Q: Do you (and your spouse or partner) currently have savings set aside for [name of child]'s postsecondary education? Types of savings include bank accounts, GICs, RESPs, RRSPs, mutual funds, investment funds, etc.
R: Yes / No

Q: How much of this training did you take through distance education such as correspondence education?
R: None of it / Less than half / About half / More than half / All of it

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 value of Registered Education Savings Plan (RESP) savings per child at the end of 2007. The numerator is an estimate of the total value of RESP savings, and the denominator is an estimate of the total number of children.

Example of Quantitative Question:

Q: How many (paid) hours did (you/your spouse or partner) usually work per week at this (most recent) job?)
R: |_|_| hours

Q: What was the value of [name of child]'s RESP at the end of 2007? Include earnings and interest as well as the Canada Education Savings Grant (CESG) and the Canada Learning Bond (CLB).
R: |_|_|_|_|_|_|_| 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 value of RESP savings at the end of 2007 saved for children whose PMK1 hope they will pursue postsecondary studies, multiply the value reported in SS_Q09P (value of RESPs at the end of 2007) by the final weight for the record, then sum this value over all records with PH_Q01P = between 4 and 12 (postsecondary).

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 value of RESPs savings at the end of 2007 for children whose PMK1 hope they will pursue postsecondary studies,

a) estimate the total value of RESPs ($\hat{X}$) as described above,
b) estimate the number of children ($\hat{Y}$) in this category by summing the final weights of all records with PH_Q01P = between 4 and 12, then
c) divide estimate a) by estimate b) ($\hat{X} / \hat{Y}$).

9.4 Guidelines for Statistical Analysis

The ASETS 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 youth respondents aged 18 to 24 is required. The steps to rescale the weights are as follows:

1) select all respondents from the file who reported DVAGE = 18 to 24;

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 DVAGE = 18 to 24;

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.

It is recommended that the bootstrap method be used to calculate variance estimates for the ASETS. The bootstrap method is a resampling method for calculating valid variance estimates for complex sampling designs and complex estimators. For the ASETS, two sets of 1,000 bootstrap weights have been produced and are provided with the data files. There is one set of bootstrap weights that should be used with the Main File, and a second set of bootstrap weights that should be used with the Course File.

Statistics Canada has developed a statistical product called BOOTVAR which estimates variances using the bootstrap method. BOOTVAR allows for the estimation of variances for totals, ratios (including proportions), differences between ratios (or proportions), linear regression models, and logistic regression models. Version 3.1 of BOOTVAR (available in SAS only) also allows for the estimation of percentiles and Chi-square tests of independence. BOOTVAR is available in both the SAS and SPSS programming languages. The most recent version of the program (SAS version 3.1 and SPSS version 3.0), user documentation, and the details of specific survey parameters can be downloaded free of charge by following these links:

SAS: http://www.statcan.gc.ca/rdc-cdr/bootvar_sas-eng.htm
SPSS: http://www.statcan.gc.ca/rdc-cdr/bootvar_spss-eng.htm

For estimates based on imputed data, it is recommended that the coefficient of variation (CV) be adjusted by an inflation factor as follows:

$$\text{CV\_adjusted} = \left(\frac{1}{\sqrt{1 - \text{ImputationRate}}} \right) \times CV \,,$$

where the CV is the coefficient of variation calculated using the bootstrap method.

9.5 Coefficient of Variation Release Guidelines

Before releasing and/or publishing any estimates from the ASETS, 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 E (or some similar identifier). They should be accompanied by a warning to caution subsequent users about the high levels of error, associated with the estimates.
3) Unacceptable	Estimates have a sample size of less than 30, or very high coefficients of variation in excess of 33.3%. Statistics Canada recommends not to release estimates of unacceptable quality. However, if the user chooses to do so then estimates should be flagged with the letter F (or some similar identifier) and the following warning should accompany the estimates: “Please be warned that these estimates [flagged with the letter F] do not meet Statistics Canada’s quality standards. Conclusions based on these data will be unreliable, and most likely invalid.”

9.6 Release Cut-off's

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 5,000 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 (nor 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.

Table of Release Cut-offs – Main File

Province	Acceptable CV 0.0% to 16.5%		Marginal CV 16.6% to 33.3%			Unacceptable CV > 33.3%	
Newfoundland and Labrador	17,500	& over	4,400	to <	17,500	under	4,400
Prince Edward Island	4,900	& over	1,200	to <	4,900	under	1,200
Nova Scotia	30,100	& over	7,600	to <	30,100	under	7,600
New Brunswick	24,300	& over	6,100	to <	24,300	under	6,100
Quebec	82,800	& over	20,600	to <	82,800	under	20,600
Ontario	95,700	& over	23,700	to <	95,700	under	23,700
Manitoba	37,200	& over	9,400	to <	37,200	under	9,400
Saskatchewan	30,300	& over	7,700	to <	30,300	under	7,700
Alberta	76,800	& over	19,200	to <	76,800	under	19,200
British Columbia	94,600	& over	23,700	to <	94,600	under	23,700
Canada	**82,000**	**& over**	**20,200**	**to <**	**82,000**	**under**	**20,200**

Table of Release Cut-offs – Course File

Province	Acceptable CV 0.0% to 16.5%		Marginal CV 16.6% to 33.3%			Unacceptable CV > 33.3%	
Newfoundland and Labrador	90,300	& over	31,700	to <	90,300	under	31,700
Prince Edward Island	18,600	& over	5,600	to <	18,600	under	5,600
Nova Scotia	113,400	& over	33,700	to <	113,400	under	33,700
New Brunswick	67,400	& over	19,400	to <	67,400	under	19,400
Quebec	394,400	& over	109,500	to <	394,400	under	109,500
Ontario	885,500	& over	238,900	to <	885,500	under	238,900
Manitoba	250,900	& over	87,700	to <	250,900	under	87,700
Saskatchewan	110,300	& over	31,600	to <	110,300	under	31,600
Alberta	360,600	& over	100,800	to <	360,600	under	100,800
British Columbia	401,300	& over	115,400	to <	401,300	under	115,400
Canada	**666,100**	**& over**	**168,600**	**to <**	**666,100**	**under**	**168,600**

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 Access and Support to Education and Training Survey (ASETS).

Main File

Province	Design Effect	Sample Size	Population
Newfoundland and Labrador	1.76	1,529	433,148
Prince Edward Island	1.73	1,468	118,756
Nova Scotia	2.00	1,814	774,025
New Brunswick	1.82	1,652	625,391
Quebec	1.94	5,530	6,524,701
Ontario	2.05	8,636	11,081,441
Manitoba	1.93	1,768	965,974
Saskatchewan	1.85	1,727	803,193
Alberta	1.91	2,749	3,090,573
British Columbia	2.05	2,882	3,719,183
Canada	**2.37**	**29,755**	**28,136,385**

Course File

Province	Design Effect	Sample Size	Population
Newfoundland and Labrador	4.71	262	227,105
Prince Edward Island	2.79	314	76,005
Nova Scotia	3.12	385	494,733
New Brunswick	2.18	332	347,374
Quebec	3.62	736	2,578,207
Ontario	6.51	1,766	7,426,596
Manitoba	7.02	395	635,376
Saskatchewan	2.61	409	581,407
Alberta	3.20	612	2,238,423
British Columbia	3.67	560	2,068,718
Canada	**6.54**	**5,771**	**16,673,943**

All coefficients of variation in the Approximate Sampling Variability Tables are approximate and, therefore, unofficial. Estimates of actual variance for specific variables may be obtained from Statistics

Canada on a cost-recovery basis, or may be calculated using the bootstrap weights provided with the data file. 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 to release 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 sub-group of the population. For example, the proportion of children whose person most knowledgeable (PMK1) have set aside savings for postsecondary education is more reliable than the estimated number of children whose PMK1 have set aside savings for postsecondary education. (Note that in the tables the coefficients of variation decline in value reading from left to right).

When the proportion or percentage is based upon the total population of the geographic area covered by the table, the CV of the proportion or percentage is the same as the CV of the numerator of the proportion or percentage. In this case, Rule 1 can be used.

When the proportion or percentage is based upon a subset of the total population (e.g. those in a particular sex or age group, reference should be made to the proportion or percentage (across the top of the table) and to the numerator of the proportion or percentage (down the left side of the table). The intersection of the appropriate row and column gives the coefficient of variation.

Rule 3: Estimates of Differences Between Aggregates or Percentages

The standard error of a difference between two estimates is approximately equal to the square root of the sum of squares of each standard error considered separately. That is, the standard error of a difference $\left(\hat{d} = \hat{X}_1 - \hat{X}_2\right)$ is:

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

where $\hat{X}_1$ is estimate 1, $\hat{X}_2$ is estimate 2, and α_1 and α_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively. The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d}$. This formula is accurate for the difference between separate and uncorrelated characteristics, but is only approximate otherwise.

Rule 4: Estimates of Ratios

In the case where the numerator is a subset of the denominator, the ratio should be converted to a percentage and Rule 2 applied. This would apply, for example, to the case where the denominator is the number of youths and the numerator is the number of youths who are postsecondary students.

In the case where the numerator is not a subset of the denominator, as for example, the ratio of the number of female postsecondary students, as compared to the number of male postsecondary students, 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 ASETS 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 4,174,026 children aged 0 to 17 have PMK1s who hope the child will pursue postsecondary studies and who have set aside (possibly with their spouse) savings for the child's postsecondary education (based on FP_Q01P or FO_Q01P). How does the user determine the coefficient of variation of this estimate?

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

2) The estimated aggregate 4,174,026 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, 2.2%.

4) The imputation rate for FP_Q01P is 1.9%. The inflation factor for imputation is approximately 1.01, and so the adjusted CV = 1.01 x 2.2% = 2.2% (see section 9.4). (The imputation rate is low enough that the CV remains the same.)

5) So the approximate coefficient of variation of the estimate is 2.2%. The finding that there were 4,174,026 (to be rounded according to the rounding guidelines in Section 9.1) children whose PMK1 hope the child will pursue postsecondary studies and have set aside savings for the child's postsecondary education is publishable with no qualifications.

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

Suppose that the user estimates that 4,174,026 / 6,721,149 = 62.1% of children aged 0 to 17 have PMK1s who hope the child will pursue postsecondary studies and who have set aside (possibly with their spouse) savings for the child's postsecondary education. How does the user determine the coefficient of variation of this estimate?

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

2) Because the estimate is a percentage which is based on a subset of the total population (i.e., children aged 0 to 17), it is necessary to use both the percentage (62.1%) and the numerator portion of the percentage (4,174,026) in determining the coefficient of variation.

3) The numerator, 4,174,026, 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. Similarly, the percentage estimate does not appear as any of the column headings, so it is necessary to use the percentage closest to it, 70.0%.

4) The figure at the intersection of the row and column used, namely 1.3% is the coefficient of variation to be used.

5) The imputation rate for FP_Q01P is 1.9%. The inflation factor for imputation is approximately 1.01, and so the adjusted CV = 1.01 x 1.3% = 1.3%. (The imputation rate is low enough that the CV remains the same.)

6) So the approximate coefficient of variation of the estimate is 1.3%. The finding that 62.1% of children have PMK1s who hope the child will pursue postsecondary studies and who have set aside savings for the child's postsecondary education can be published with no qualifications.

Example 2B: Estimates of Proportions or Percentages of Job-related Training Activities Possessing a Characteristic

Suppose that the user estimates that 1,162,711 / 15,462,178 = 7.5% of job-related courses, workshops or seminars taken by persons aged 18 to 64 between July 2007 and June 2008, were taken entirely through distance education. How does the user determine the coefficient of variation of this estimate?

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

2) Because the estimate is a percentage which is based on a subset of the total population (i.e., job-related courses, workshops or seminars taken by persons aged 18 to 64 between July 2007 and June 2008), it is necessary to use both the

percentage (7.5%) and the numerator portion of the percentage (1,162,711) in determining the coefficient of variation.

3) The numerator, 1,162,711, 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 1,000,000. Similarly, the percentage estimate does not appear as any of the column headings, so it is necessary to use the percentage closest to it, 10.0%.

4) The figure at the intersection of the row and column used, namely 13.0% is the coefficient of variation to be used.

5) Variable CN_Q04 was not imputed, so the CV does not need to be adjusted.

6) So the approximate coefficient of variation of the estimate is 13.0%. The finding that 7.5% of job-related courses, workshops or seminars were taken entirely through distance education can be published with no qualifications.

Example 3: Estimates of Differences Between Aggregates or Percentages

Suppose that the user estimates that 2,890,532 / 4,593,340 = 62.9% of young children aged 0 to 12 have PMK1s who hope the child will pursue postsecondary studies and who have set aside (possibly with their spouse) savings for the child's postsecondary education. Suppose the user also estimates that the equivalent proportion for older children aged 13 to 17 is 1,283,494 / 2,127,809 = 60.3%. How does the user determine the coefficient of variation of the difference between these two estimates?

1) Using CANADA coefficient of variation table for the Main File in the same manner as described in Example 2A gives the CV of the estimate for younger children as 1.5%, and the CV of the estimate for older children as 2.1%.

2) The imputation rates for FP_Q01P are 2.45% for younger children and 0.74% for older children. The imputation rates are low enough that the adjusted CVs remain the same.

3) 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 (younger children), $\hat{X}_2$ is estimate 2 (older children), 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 $\hat{d} = 0.629 - 0.603 = 0.026$ is:

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{\left[(0.629)(0.015)\right]^2 + \left[(0.603)(0.021)\right]^2} \\ &= \sqrt{(0.000089) + (0.000160)} \\ &= 0.016\end{aligned}$$

4) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d} =$ 0.016 / 0.026 = 0.615.

5) So the approximate coefficient of variation of the difference between the estimates is 61.5%. 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 F (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 1,060,647 female youths aged 18 to 24 have at least some postsecondary education, while 1,004,179 male youths have at least some postsecondary education. The user is interested in comparing the estimate of females versus males 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 female youths who have at least some postsecondary education. The denominator of the estimate ($\hat{X}_2$) is the number of male youths who have at least some postsecondary education.

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

3) The numerator of this ratio estimate is 1,060,647. 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, 4.6%.

4) The denominator of this ratio estimate is 1,004,179. 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, 4.6%

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:

$$\begin{aligned}\alpha_{\hat{R}} &= \sqrt{(0.046)^2 + (0.046)^2} \\ &= \sqrt{0.00212 + 0.00212} \\ &= 0.0651\end{aligned}$$

6) The imputation rate for EA_Q01 for youths is 11.1%. The inflation factor for imputation is approximately 1.06, and so the adjusted CV = 1.06 x 6.5% = 6.9%.

7) The obtained ratio of female youths versus male youths who have at least some postsecondary education is 1,060,647 / 1,004,179 which is 1.06 (to be rounded according to the rounding guidelines in Section 9.1). The coefficient of variation of this estimate is 6.9%, which makes the estimate releasable with no qualifications.

Example 5: Estimates of Differences of Ratios

Suppose that the user estimates that the ratio of female youths with at least some postsecondary education, to male youths with at least some postsecondary education is 1.11 for Ontario while it is 0.97 for British Columbia. 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 Ontario ratio ($\hat{R}_1$) and the British Columbia ratio ($\hat{R}_2$) as in Example 4. The approximate CV for the Ontario ratio is 10.8% and 20.1% for British Columbia.

2) The imputation rates for EA_Q01 are 13.1% for Ontario and 13.7% for British Columbia. The adjusted CVs are 11.6% for Ontario and 21.7% for British Columbia.

3) 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 $\hat{d}$ = 1.11 – 0.97 = 0.14 is:

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{\left[(1.11)(0.116)\right]^2 + \left[(0.97)(0.217)\right]^2} \\ &= \sqrt{(0.0166) + (0.0443)} \\ &= 0.247\end{aligned}$$

4) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d}$ = 0.247 / 0.14 = 1.76.

5) So the approximate coefficient of variation of the difference between the estimates is 176%. 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 F (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 children aged 0 to 17 with PMK1s who hope the child will pursue postsecondary studies and who have set aside (possibly with their spouse) savings for the child's postsecondary education (from Example 2A, Section 10.1.1) would be calculated as follows:

$\hat{X}$ = 62.1% (or expressed as a proportion 0.621)

t = 2

$\alpha_{\hat{x}}$ = 1.3% (0.013 expressed as a proportion) is the coefficient of variation of this estimate as determined from the tables.

$CI_{\hat{x}}$ = {0.621– (2) (0.621) (0. 013), 0.621+ (2) (0.621) (0. 013)}

$CI_{\hat{x}}$ = {0.621 – 0.016, 0.621 + 0.016}

$CI_{\hat{x}}$ = {0.605, 0.637}

With 95% confidence it can be said that between 60.5% and 63.7% of children aged 0 to 17 have PMK1s who hope the child will pursue postsecondary studies and who have set aside (possibly with their spouse) savings for the child's postsecondary education.

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 $t = \dfrac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}}$ is between -2 and 2, then no conclusion about the difference between the characteristics is justified at the 5% level of significance. If however, this ratio is smaller than -2 or larger than +2, the observed difference is significant at the 0.05 level. That is to say that the difference between the estimates is significant.

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 younger children aged 0 to 12 versus older children aged 13 to 17 with PMK1s who hope the child will pursue postsecondary studies and who have set aside (possibly with their spouse) savings for the child's postsecondary education. From Example 3, Section 10.1.1, the standard error of the difference between these two estimates was found to be 0.016. Hence,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0.629 - 0.603}{0.016} = \frac{0.026}{0.016} = 1.63$$

Since t = 1.63 is less than 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 ASETS are primarily categorical in nature, this has not been done.

As a general rule, however, the coefficient of variation of a quantitative total will be larger than the coefficient of variation of the corresponding category estimate (i.e., the estimate of the number of persons contributing to the quantitative estimate). If the corresponding category estimate is not releasable, the quantitative estimate will not be either. For example, the coefficient of variation of the total value of RESPs saved for children would be greater than the coefficient of variation of the corresponding proportion of children with RESPs. Hence, if the coefficient of variation of the proportion is unacceptable (making the proportion not releasable), then the coefficient of variation of the corresponding quantitative estimate will also be unacceptable (making the quantitative estimate not releasable).

Coefficients of variation of such estimates can be derived using the bootstrap weights provided with the data file.

10.5 Coefficient of Variation Tables

Refer to Appendix G for the coefficient of variation tables for the Main File.

Refer to Appendix H for the coefficient of variation tables for the Course File.

11.0 Weighting

This chapter outlines the weighting steps that were performed to derive the final weights for the Access and Support to Education and Training Survey (ASETS). The records were first categorized as follows:

Category		Description	Count
Overlap	OVLP	Overlap with other surveys – not sent to data collection	3,503
Unresolved	UNR1	No screener and no roster; unresolved outcome code	576
	UNR2	No screener; in-scope outcome code	12,127
	UNR3	In-scope screener but no household roster	1,528
Out-of-scope	OOS1	Out-of-scope outcome code, except those in OOS2	15,225
	OOS2	Cell phones or secondary dwellings with completed screener	3,069
	OOS3	Completed roster, out-of-scope household (all members 65 and over)	2,204
In-scope	SHR	Sharers: respondents who gave permission to share their data	29,755
	NSHR	Non-sharers: respondents who did not give permission to share their data	1,754
	NR	Non-respondents: completed roster but no data for selected respondent	5,320
Total initial sample size			**75,061**

The following weighting steps were performed:

1. Calculate design weights

Each of the 75,061 records in the initial sample was assigned a design weight, W_1. For the Census portion of the sample, the design weights were equal to the Census 2B weight multiplied by the inverse of the probability of selection. For the Administrative stratum, the design weights were equal to the inverse of the probability of selection.

2. Adjust for overlap with other surveys (OVLP)

Weighting adjustments were performed to take into account 3,503 telephone numbers which were not sent to the field for data collection because they overlapped with telephone numbers already interviewed for other surveys. Weighting classes were formed based on stratum, and Census data for some records. The following adjustments were calculated within each weighting class:

$$W_2 = \left(\frac{\sum W_1 \text{ for records sent to the field} + \sum W_1 \text{ for records in OVLP}}{\sum W_1 \text{ for records sent to the field}} \right) \times W_1$$

3. Adjust for unresolved units with no data (UNR1)

Weighting adjustments were performed to take into account 576 records that were sent to the field for data collection but no data were received to determine whether they are in-scope for the survey. The following adjustments were calculated within each stratum:

$$W_3 = \left(\frac{\sum W_2 \text{ for records sent to the field}}{\sum W_2 \text{ for records classified as SHR, NSHR, NR, OOS1, OOS2, OOS3, UNR2, UNR3}} \right) \times W_2$$

4. Adjust for unresolved units with no screener (UNR2)

Weighting adjustments were performed to take into account 12,127 records which have an in-scope outcome code but no screener and no household roster. Propensity to respond was modelled using Census data (for the Census portion of the sample) and paradata. Weighting classes were formed based on the predicted propensity to respond. The following adjustments were calculated within each weighting class:

$$W_4 = \left(\frac{\sum W_3 \text{ for records classified as SHR, NSHR, NR, OOS2, OOS3, UNR3} + \sum W_3 \text{ for records classified as UNR2}}{\sum W_3 \text{ for records classified as SHR, NSHR, NR, OOS2, OOS3, UNR3}} \right) \times W_3$$

5. Adjust for unresolved units with no roster (UNR3)

Weighting adjustments were performed to take into account 1,528 records which have an in-scope outcome code and a completed screener but no household roster. Propensity to respond was modelled using Census data (for the Census portion of the sample) and paradata. Weighting classes were formed based on the predicted propensity to respond. The following adjustments were calculated within each weighting class:

$$W_5 = \left(\frac{\sum W_4 \text{ for records classified as SHR, NSHR, NR, OOS3} + \sum W_4 \text{ for records classified as UNR3}}{\sum W_4 \text{ for records classified as SHR, NSHR, NR, OOS3}} \right) \times W_4$$

6. Calculate person level weights

Person level weights were calculated for the in-scope cases with a completed screener and roster, and a selected household member. The weights from the previous step were multiplied by the inverse of the probability of selection for the selected person. A cap was placed on the adjustment in order to avoid extreme weights. The person level weights were denoted W_6.

7. Adjust for person non-response (NR)

Weighting adjustments were performed to take into account 5,320 in-scope cases where the household member selected from the roster did not provide sufficient data to be considered a respondent. Propensity to respond was modelled using Census data (for the Census portion of the sample) and the ASETS roster data. Weighting classes were formed based on the predicted propensity to respond. The following adjustments were calculated within each weighting class:

$$W_7 = \left(\frac{\sum W_6 \text{ for records classified as SHR, NSHR} + \sum W_6 \text{ for records classified as NR}}{\sum W_6 \text{ for records classified as SHR, NSHR}} \right) \times W_6$$

8. Calibration for the Master File (SHR and NSHR)

The weights of the 31,509 respondents on the master file (sharers and non-sharers) were calibrated so that the sum of the weights are equal to July 2008 demographic counts, at the province by age group by gender level. Calibration was also used to reduce the under-representation of youths aged 18 to 24 who do not live with their parents. The final (person level) weights for the master file are called WTPM.

9. Calibration for the Share File (SHR)

The weights of the 29,755 respondents on the share file (respondents who agreed to share) were calibrated to the same province by age group by gender counts as the master file. In addition, the weights were calibrated so that the share file produces some of the same estimates as the master file. The final (person level) weights for the share file are called WTPS.

10. Calculation of Course Level Weights for the Master Course File

Course level weights were calculated for all respondents who reported having taken training for job or career reasons in CR1_Q02. For respondents who reported ten courses or less in EC_Q03, the course weights were calculated as follows:

WTCM = WTPM x Number of Job-related Courses

For respondents who reported more than ten courses in EC_Q03, the course weights were calculated as follows:

WTCM = WTPM x (EC_Q03 / 10) x Number of Job-related Courses

The course level weights should be used with the course level data file, which contains the data for the randomly selected course. The weights for the master course file are called WTCM.

11. Calculation of Course Level Weights for the Share Course File

Course level weights were calculated for all respondents who reported having taken training for job or career reasons in CR1_Q02. For respondents who reported ten courses or less in EC_Q03, the course weights were calculated as follows:

WTCS = WTPS x Number of Job-related Courses

For respondents who reported more than ten courses in EC_Q03, the course weights were calculated as follows:

WTCS = WTPS x (EC_Q03 / 10) x Number of Job-related Courses

The course level weights should be used with the course level data file, which contains the data for the randomly selected course. The weights for the share course file are called WTCS.

12.0 Questionnaires

The Access and Support to Education and Training Survey (ASETS) questionnaire was used in 2008 to collect the information for the survey. The file ASETS2008_QuestE.pdf contains the English questionnaire.

13.0 Record Layout with Univariate Frequencies

Refer to ASETS2008_Main_CdBk.pdf for the record layout with univariate frequencies for the Main File.

Refer to ASETS2008_Course_CdBk.pdf for the record layout with univariate frequencies for the Course File.