Microdata User Guide

Survey of Approaches to Educational Planning

1999

Statistics Canada Statistique Canada

Canada

Table of Contents

1.0 Introduction

The Survey of Approaches to Educational Planning (SAEP) was conducted by Statistics Canada in October 1999 with the cooperation and support of Human Resources Development Canada. This manual has been produced to facilitate the manipulation of the microdata file of the survey results.

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

Statistics Canada

Client Services
Centre for Education Statistics
Telephone: (613) 951-7608 or call toll-free 1 800 307-3382
Fax: (613) 951-9040
E-mail: educationstats@statcan.ca

2.0 Background

Statistics Canada was approached by Human Resources Development Canada (HRDC) to conduct a survey to examine how Canadians are preparing their children for post-secondary education. The objective was to examine both the financial and "cultural" aspects of preparation. Financial preparation includes parents' expectations of the cost of their children's post-secondary education as well as the amount and type of savings that have been made for their education. Cultural aspects include parents' involvement in their children's schooling and the provision of educational resources.

Saving for post-secondary education is not an activity that is restricted to parents of school-aged children. Many other individuals, such as grandparents, also save for a child's post-secondary education. In order to obtain educational savings data for all households in Canada, HRDC opted to include households without children in the survey as well. These households were asked financial savings questions only. The inclusion of these households allows statements to be made about all educational savings in Canada, not just savings by households with children.[1]

The 1999 Survey of Approaches to Educational Planning is the first survey conducted by Statistics Canada to collect detailed information about how Canadians prepare their children for post-secondary education. Given the growing importance of post-secondary education in Canada and recent increases in the costs, the survey will provide a key first look at how Canadians are getting ready for their children's post-secondary education.

[1] Data for households with children are not available on the public use microdata file, but can be obtained by custom tabulations from Client Services, Centre for Education Statistics.

3.0 Objectives

The objective of the Survey of Approaches to Educational Planning is to profile two ways Canadians can prepare for their children's post-secondary education:

1) Non-financial preparation - how are parents involved in their children's current schooling and learning?

2) Financial preparation - what is the awareness of the cost of post-secondary education, what are the plans for paying for children's education, and what are the past and current savings?

4.0 Concepts and Definitions

This chapter outlines concepts and definitions of interest to the users. The concepts and definitions used in the Labour Force Survey (LFS) are described in Section 4.1 while those specific to the Survey of Approaches to Educational Planning (SAEP) are given in Section 4.2. Users are referred to Chapter 12.0 of this document for a copy of the actual survey forms used.

4.1 Labour Force Survey Concepts and Definitions

Labour Force Status

Designates the status of the respondent vis-à-vis the labour market: a member of the non-institutional population 15 years of age and over is either **employed, unemployed** or **not in the labour force**.

Employment

Employed persons are those who, during the reference week:

a) did any work[2] at all at a job or business; or
b) had a job but were not at work due to factors such as own illness or disability, personal or family responsibilities, vacation, labour dispute or other reasons (excluding persons on layoff, between casual jobs, and those with a job to start at a future date).

Unemployment

Unemployed persons are those who, during the reference week:

a) were on temporary layoff during the reference week with an expectation of recall and were available for work; or
b) were without work, had actively looked for work in the past four weeks, and were available for work;[3] or
c) had a new job to start within four weeks from the reference week, and were available for work.

[2] Work includes any work for pay or profit, that is, paid work in the context of an employer-employee relationship, or self-employment. It also includes unpaid family work, which is defined as unpaid work contributing directly to the operation of a farm, business or professional practice owned and operated by a related member of the same household. Such activities may include keeping books, selling products, waiting on tables, and so on. Tasks such as housework or maintenance of the home are not considered unpaid family work.

[3] Persons are regarded as available for work if they:

i) reported that they could have worked in the reference week if a suitable job had been offered; or if the reason they could not take a job was of a temporary nature such as: because of own illness or disability, personal or family responsibilities, because they already have a job to start in the near future, or because of vacation (prior to 1997, those on vacation were not considered available).
ii) were full-time students seeking part-time work who also met condition i) above. Full-time students currently attending school and looking for full-time work are not considered to be available for work during the reference week.

Not in the Labour Force

Persons not in the labour force are those who, during the reference week, were unwilling or unable to offer or supply labour services under conditions existing in their labour markets, that is, they were neither employed nor unemployed.

Industry and Occupation

The Labour Force Survey provides information about the occupation and industry attachment of employed and unemployed persons, and of persons not in the labour force who have held a job in the past 12 months. Since 1997, these statistics have been based on the North American Industry Classification System (NAICS) and the Standard Occupational Classification (SOC-91). Prior to 1997, the 1980 Standard Industrial Classification and the 1980 Standard Occupational Classification were used.

Reference Week

The entire calendar week (from Sunday to Saturday) covered by the Labour Force Survey each month. It is usually the week containing the 15th day of the month. The interviews are conducted during the following week, called the Survey Week, and the labour force status determined is that of the reference week.

Full-time Employment

Full-time employment consists of persons who usually work 30 hours or more per week at their main or only job.

Part-time Employment

Part-time employment consists of persons who usually work less than 30 hours per week at their main or only job.

4.2 Survey of Approaches to Educational Planning Concepts and Definitions

Education costs

Tuition and other related education costs (books, co-op program fees, lab supplies, etc.).

Government student loans

Loans subsidized by federal and/or provincial governments that are used to pay for post-secondary studies.

Grants or bursaries

A sum of money given to a person by an organization to pay for the person to study. Grants and bursaries do not have to be repaid by the recipient.

In-trust for account

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.

Living expenses

Any expenses that a student would incur other than direct education costs. The major living expense would likely be rent/accommodation. Other expenses could include transportation, food, entertainment, clothing, etc.

Non-student loans from a financial institution

Includes personal loans, personal lines of credit and first or second mortgages.

Other loans

Includes loans from employers, salary advances and credit cards.

Post-secondary education

Any type of formal education after high school including college and university as well as apprenticeships, trade-vocational programs, Collèges d'enseignement général et professionnel (CEGEP in Quebec) and other programs.

Registered Education Savings Plan (RESP)

A tax-sheltered means of saving to finance a child's post-secondary education. Once the child starts post-secondary studies, the earnings in the plan are used to pay for tuition or other expenses related to his/her studies.

RESP – Group plans

RESPs that operate on a pooling principle where the plan redistributes the earnings among all the beneficiaries in the plan.

RESP – Individual plan

RESPs are sold by vendors that separate each RESP contract from the other independent RESP contracts that they manage.

Registered Retirement Savings Plan (RRSP)

A capital accumulation program designed to encourage saving for retirement. Contributions are tax-deductible within certain limits.

5.0 Survey Methodology

The Survey of Approaches to Educational Planning (SAEP) was administered in October 1999 to a sub-sample of the dwellings in the Labour Force Survey (LFS) sample, and therefore its sample design is closely tied to that of the LFS. The LFS design is briefly described in Sections 5.1 to 5.4.[4] Sections 5.5 and 5.6 describe how the SAEP departed from the basic LFS design in October 1999.

5.1 Population Coverage

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

5.2 Sample Design

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

5.2.1 Primary Stratification

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

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

5.2.2 Types of Areas

The primary strata (ER/EIER intersections) are further disaggregated into three types of areas: rural, urban, and remote areas. Urban and rural areas are loosely based on the Census definitions of urban and rural, with some exceptions to allow for the formation of strata in some areas. Urban areas include the largest CMAs down to the smallest villages categorized by the 1991 Census as urban (1,000 people or more), while rural areas are made up of areas not designated as urban or remote.

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

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

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

5.2.3 Secondary Stratification

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

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

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

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

5.2.4 Cluster Delineation and Selection

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

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

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

Within each of the secondary strata in rural areas, where necessary, further stratification is carried out in order to reflect the differences among a number of socio-economic

characteristics within each stratum. Within each rural stratum, six EAs or two or three groups of EAs are sampled as clusters.

5.2.5 Dwelling Selection

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

5.2.6 Person Selection

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

5.3 Sample Size

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

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

5.4 Sample Rotation

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

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

5.5 Modifications to the Labour Force Survey Design for the Survey of Approaches to Educational Planning

The Survey of Approaches to Educational Planning used five of the six rotation groups in the October 1999 LFS sample. Unlike the LFS where information is collected for all eligible household members, the SAEP only collected information from one household member about the children 18 years of age or younger in the household and the savings activities of the household.

Upon completion of the Labour Force Survey interview, the interviewers asked to speak to the Survey of Approaches to Educational Planning respondent. If there were no children 18 years of age or younger in the household, the SAEP respondent was the same as the LFS respondent. In households that had children 18 years of age or younger, the SAEP respondent was the person most knowledgeable about the children in the household and about any plans made for the children's post-secondary education.

5.6 Sample Size by Province for the Survey of Approaches to Educational Planning

The following table shows the number of households and children in the LFS sampled rotations that were eligible for the Survey of Approaches to Educational Planning supplement.

Province	Household Sample Size *	Child Sample Size **
Newfoundland and Labrador	1,465	965
Prince Edward Island	1,069	670
Nova Scotia	2,650	1,557
New Brunswick	2,300	1,466
Quebec	8,261	4,819
Ontario	12,404	8,143
Manitoba	3,071	1,929
Saskatchewan	3,238	2,211
Alberta	3,261	2,392
British Columbia	3,880	2,314
Canada	**41,599**	**26,466**

* Only LFS households in rotation groups 1, 2, 3, 5 and 6 who had given permission to be contacted by telephone were eligible for the SAEP supplement.

** Data was collected for all children within those households eligible for the SAEP supplement, up to a maximum of three children per household. In households with more than three children, the children to be sampled were randomly selected.

6.0 Data Collection

Data collection for the Labour Force Survey (LFS) is carried out each month during the week following the LFS reference week. The reference week is normally the week containing the 15th day of the month.

6.1 Interviewing for the Labour Force Survey

Statistics Canada interviewers are employees hired and trained to carry out the LFS and other household surveys. Each month they contact the sampled dwellings to obtain the required labour force information. Each interviewer contacts approximately 75 dwellings per month.

Dwellings new to the sample are usually contacted through a personal visit using the computer-assisted personal interview (CAPI). The interviewer first obtains socio-demographic information for each household member and then obtains labour force information for all members aged 15 and over who are not members of the regular armed forces. Provided there is a telephone in the dwelling and permission has been granted, subsequent interviews are conducted by telephone. This is done out of a centralized computer-assisted telephone interviewing (CATI) unit where cases are assigned randomly to interviewers. As a result, approximately 85% of all households are interviewed by telephone. In these subsequent monthly interviews, the interviewer confirms the socio-demographic information collected in the first month and collects the labour force information for the current month.

In each dwelling, information about all household members is usually obtained from one knowledgeable household member. Such “proxy” reporting, which accounts for approximately 65% of the information collected, is used to avoid the high cost and extended time requirements that would be involved in repeat visits or calls necessary to obtain information directly from each respondent.

If, during the course of the six months that a dwelling normally remains in the sample, an entire household moves out and is replaced by a new household, information is obtained about the new household for the remainder of the six-month period.

At the conclusion of the LFS monthly interviews, interviewers introduce the supplementary survey, if any, to be administered to some or all household members that month.

6.2 Supervision and Quality Control

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

6.3 Non-response to the Labour Force Survey

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

Each month, after all attempts to obtain interviews have been made, a small number of non-responding households remain. For households non-responding to the LFS and for which LFS information was obtained in the previous month, this information is brought forward and used as the current month's LFS information. No supplementary survey information is collected for these households.

6.4 Data Collection Modifications for the Survey of Approaches to Educational Planning

Due to operational constraints during the time of collection, the Survey of Approaches to Educational Planning (SAEP) was conducted as a paper and pencil survey. Labels identifying the selected individuals were produced at head office prior to survey week and then attached to the questionnaires in the Statistics Canada regional offices. If the SAEP respondent was not the same as the LFS respondent, the interviewer indicated this on the questionnaire.

The SAEP was administered to one individual per household. If there were no children 18 years of age or younger in the household, the SAEP respondent was the same as the LFS respondent. In households that had children 18 years of age or younger, the SAEP respondent was the person most knowledgeable about the children in the household and about any plans made for the children's post-secondary education.

Upon completion of the Labour Force Survey interview, the interviewer asked to speak to the SAEP respondent if it was not the same as the LFS respondent. If the SAEP respondent was not available, the interviewer arranged for a convenient time to phone back. The collection period was extended by one week to allow the interviewers time to contact the SAEP respondents.

The SAEP collected information about individual children in the household to a maximum of three children. In households with more than three children 18 years of age or younger, the interviewers were required to manually select three children using a grid on the questionnaire label.

6.5 Non-response to the Survey of Approaches to Educational Planning

For households responding to the LFS, the next stage of data collection was to administer the Survey of Approaches to Educational Planning. In total, 41,599 households were eligible for the supplementary survey. The SAEP interview was completed for 36,130 of these households for a response rate of 86.9%. In the eligible households, there were 26,466 children 18 years of age or younger. Interviewers collected information for 20,353 of these children for a response rate of 76.9%. More detailed information on response rates is presented in Chapter 8.0 (Data Quality).

7.0 Data Processing

The main output of the Survey of Approaches to Educational Planning (SAEP) is a "clean" microdata file. This chapter presents a brief summary of the processing steps involved in producing this file.

7.1 Data Capture

Capture of survey data was accomplished using the data capture facilities located in each of Statistics Canada's Regional Offices. During this process any document containing at least one interviewer-completed item was captured and an unedited version of the computer record was electronically transmitted to Ottawa for further processing. In total 42,425 documents were captured and transmitted for the survey

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.

Edits were also performed to ensure that the number, age and sex of the children in the household collected by the SAEP matched those of the Labour Force Survey (LFS).

7.3 Coding of Open-ended Questions

A few data items on the questionnaire were recorded by interviewers in an open-ended format. A total of seven partially or completely open-ended questions were included in the survey. These were items relating to the reasons a child did not attend school in 1998-1999, ethnicity and languages spoken in the household.

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 questions. 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 questions 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 1999 SAEP, donor imputation was used to fill missing data for six key items due to item non-response. Further information on the imputation process is given in Chapter 8.0 (Data Quality).

7.5 Creation of Derived Variables

A number of data items on the microdata file have been derived by combining items on the questionnaire in order to facilitate data analysis. The following are some examples of the derived variables that have been created.

Child Derived Variables

Child Public Use Microdata File and Child Master file:

Household savings status (HSAVST)
- Saving, children 0 to 18 in household
- Not saving, children 0 to 18 in household

Status of children for whom there are savings in the household (HSAVST2)
- Saving for children who live in the household only
- Saving for children who live outside the household only
- Saving for children who live in and outside the household
- Not saving

Household Registered Education Savings Plan (RESP) status (HRESPST)
- Saving household using RESPs
- Saving household not using RESPs
- No savings

Total number of children living inside and outside the household for whom there are savings (TCHSAVAL)

Total number of children living in the household for whom there are savings (TCHSAVIN)

Child Public Use Microdata File only:

Main language spoken in household (MAINLANG)
- English only
- French only
- Other only
- More than one language

Other language(s) spoken in household (OTHLANG)
- No other language spoken in household
- Other language(s) spoken in household

Household income (INCOMEDV)
- $0 to $9,999
- $10,000 to $14,999
- $15,000 to $19,999
- $20,000 to $29,999
- $30,000 to $39,999
- $40,000 to $49,999

- $50,000 to $59,999
- $60,000 to $79,999
- $80,000 or more

Child Master File only:

Total household savings in 1998 (HSAV1998)

Total household savings in 1998 for children living in the household only (HSAV98IN)

Total household savings in 1999 (HSAV1999)

Total household savings in 1999 for children living in the household only (HSAV99IN)

Total household savings planned for the rest of 1999 (HSAVRT99)

Total household savings planned for the rest of 1999 for children living in the household only (HSAVR99I)

Total household savings (HSAVTOT)

Total household savings for children living in the household only (HSAVTOTI)

Total RESP savings (HRESPSAV)

Total RESP savings for children living in the household only (HRESPSIN)

Main language spoken in household (MAINLANG)
- English only
- French only
- Other only
- English and French only
- English and other
- French and other
- English, French and other

Household income (INCOMEDV)
- No income
- $1 to $4,999
- $5,000 to $9,999
- $10,000 to $14,999
- $15,000 to $19,999
- $20,000 to $29,999
- $30,000 to $39,999
- $40,000 to $49,999
- $50,000 to $59,999
- $60,000 to $79,999
- $80,000 or more

There are also a number of derived variables that were created at the request of Human Resources Development Canada (HRDC) and are available on the master and shared microdata files. The first type of derived variable created determines if an equal amount of money has been saved for all children in a household. For example, in a household with three children, the respondent reported that in 1998 household members had saved $500 for the first child, $500 for the second child and $500 for the third child. This type of derived variable was created for:

- Amounts saved in 1998 (DV32)
- Amounts saved in 1999 (DV33)
- Total savings (DV35)
- Savings in RESPs (DV40)

The second type of variable was created to flag child records with inconsistent saving amounts. HRDC requested that these inconsistencies be left on the master and shared microdata files. This type of derived variable was created to check if:

- The total amount of money saved for a child is less than or equal to the sum of the amounts saved in 1998 and 1999.
- The amount of money expected to be saved by the time the child starts his/her post-secondary education is less than or equal to the total amount of money already saved.
- The total amount of money saved for a child is less than or equal to the amount saved in RESPs.

7.6 Weighting

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

The weighting phase is a step which calculates, for each record, what this number is. This weight appears on the microdata file, and **must** be used to derive meaningful estimates from the survey. For example, if the number of children who have savings for their post-secondary education 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.

7.7 Suppression of Confidential Information

It should be noted that the “Public Use” microdata files described above differ in a number of important respects from the survey “master” files held by Statistics Canada. These differences are the result of actions taken to protect the anonymity of individual survey respondents. Users requiring access to information excluded from the microdata files may purchase custom tabulations. Estimates generated will be released to the user, subject to meeting the guidelines for analysis and release outlined in Chapter 9.0 of this document.

Geographic Identifiers

The survey master data file includes explicit geographic identifiers for region, province, economic region and census metropolitan area. It is also possible to obtain, where sample sizes permit, estimates by urban size class. The survey public use microdata files do not contain any geographic identifiers below the regional level (Atlantic, Quebec, Ontario, Prairies, British Columbia).

Other Variables - Child Public Use Microdata File only

For confidentiality reasons, the values of a number of variables were grouped or the top and bottom values of the variables were capped. Some examples of these actions are listed below.

AGE, FAGE and SAGE
The ages of the child, the first parent and the second parent are grouped into ranges on the public use microdata file.

FIR_EDUC and SEC_EDUC
The highest level of education of the first parent and the second parent of the child are grouped into five categories on the public use microdata file

A1
The number of children 18 years of age or younger living in the household is capped at four or more children on the public use microdata file.

Q1
The relationship of the child to the person providing the information about the child is grouped into three categories on the public use microdata file.

ORIGIN
The ethnic background of the parents of the children in the household is suppressed on the public use microdata file. Instead, a variable that combines the ethnic background of the children's parents (ORIGIN) is contained on the public use microdata file.

MAINLANG and OTHLANG
The main language spoken in the household (MAINLANG) is grouped into four categories on the public use microdata file. Details about other languages spoken in the household are suppressed. In its place, a variable is provided on the microdata file that indicates if another language is spoken in the household (OTHLANG).

G9FGH, G9KL and G9MN
On the public use microdata file, some sources of income are grouped together.

8.0 Data Quality

8.1 Response Rates

The following table summarizes the response rates to the Labour Force Survey (LFS) and to the Survey of Approaches to Educational Planning (SAEP).

Province	Household Response Rate for Full LFS* October 1999	Household Response Rate for LFS Rotations 1, 2, 3, 5 and 6	SAEP Household Response Rate **	Number of Children 0 to 18 Years of Age of Respondents to SAEP	SAEP Child Response Rate ***
	%	%	%		%
Newfoundland and Labrador	92.6	93.1	91.5	828	85.8
Prince Edward Island	94.9	96.1	89.4	526	78.5
Nova Scotia	90.7	91.8	87.6	1,197	76.9
New Brunswick	90.0	90.6	86.9	1,147	78.2
Quebec	94.5	95.6	90.4	3,942	81.8
Ontario	96.1	96.9	86.0	6,161	75.7
Manitoba	94.8	96.4	86.3	1,477	76.6
Saskatchewan	94.6	95.2	85.0	1,638	74.1
Alberta	95.1	95.7	84.1	1,740	72.7
British Columbia	95.1	95.5	83.3	1,697	73.3
Canada	**94.5**	**95.4**	**86.9**	**20,353**	**76.9**

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

** The SAEP response rate is the number of households responding to the SAEP as a percentage of the number of households who were eligible for the SAEP. (Only LFS households in rotation groups 1, 2, 3, 5 and 6 who had given permission to be contacted by telephone were eligible for the SAEP.)

*** The SAEP overall response rate is the number of children for whom a response to the SAEP supplement was received as a percentage of the number of children who were eligible in those households selected for the SAEP. A maximum of three children per household could be selected. Response rates were significantly lower in the 16 to 18 year old age category at 69.7%.

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 was made to reduce non-sampling errors in the survey. Quality assurance measures were implemented at each step of the data collection and processing cycle to monitor the quality of the data. These measures 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

Because the SAEP was a supplement to the LFS, the frame used was the LFS frame. Any non-response to the LFS had an impact on the SAEP frame. The quality of the sampling variables in the frame was very high. The SAEP sample consisted of five rotation groups from the LFS. The criteria used for the SAEP selection (like rotation group) were not missing for any LFS record.

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

All variables on the LFS frame are updated monthly.

8.2.2 Data Collection

Interviewer training consisted of reading the SAEP Supervisor's Manual, Procedures Manual and Interviewer's Manual and discussing any questions with senior interviewers before the start of the survey. A description of the background and objectives of the survey was provided, as well as a glossary of terms and a set of questions and answers. Interviewers collected the SAEP information after the LFS information was collected.

An introductory letter was mailed to respondents from the regional offices 10 days before data collection began. The main purpose of the letter was to improve the quality of the data collected. The letter advised respondents that the survey would be collecting data about post-secondary savings. Respondents would then be able to look up any relevant savings information, if required, before being contacted for the survey.

8.2.3 Data Processing

During data capture, key variables were verified 100% to ensure data quality. The key variables were the number of children in the household, the age and sex of the children, the age of the child when savings were first started, types of savings plans, all savings amounts, income and permission to share and link the data.

During processing of the data, 504 SAEP records did not match to the list of labels created at head office. It was determined that most of these extra records were from LFS households that had not given permission to call. However, interviewers had been able to obtain interviews with these households through personal interviews. It was decided to retain these records.

Data processing of the SAEP 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 households 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.

For the 1999 SAEP, donor imputation was used to fill missing data in household income and six key items. This was done in order to provide complete data, thereby allowing for totals to be estimated (e.g., total Registered Education Savings Plan (RESP) contributions in Ontario).

The six key items gathered information on the current value of or the annual contribution to, savings for the post-secondary education of children aged 0 to18 years. The savings were in terms of RESPs or other savings (e.g., term deposits, guaranteed investment certificates (GIC), savings bonds, Registered Retirement Savings Plans (RRSP), mutual funds). The corresponding modules, item labels, and item descriptions are given in the table below.

Module	Item Label	Item Description
G - General information	G10C	Household income
B, C and D, - Savings for post-secondary education	Q32A	Contributions to savings in 1998
	Q33A	Contributions to savings so far in 1999
	Q34A	Expected contributions to savings for the rest of 1999
	Q35A	Current value of savings
	Q40A	Current value of RESP savings
	Q43A	Current value of savings by others

Note: Any related derived variables to these variables are based on imputed data as well. For example, the derived variable INCOMEDV on the child public use microdata file was derived from the imputed variable G10C.

Because the six items depend on previous questions (lead-ins), missing values in the lead-ins were imputed first. The lead-ins ask whether there are (or will be) savings and, if so, whether these savings are for the post-secondary education of children aged 0 to 18 years. The corresponding modules, lead-in labels, and lead-in descriptions are given in the table below.

Module	Lead-in	Lead-in description	Lead-in to:
B, C and D - Planning for post-secondary education	Q20	Have parents/guardians ever saved for child's post-secondary education?	Q32A, Q33A, Q34A, Q35A
	Q37A	Are parents using RESPs to save for child's post-secondary education?	Q40A
	Q42	Do others have savings for child's post-secondary education?	Q43A

Imputation involved filling the missing values in household income, the six items and/or the lead-ins on a given record (the "recipient" record) using another record whose values were all known and whose characteristics were the "closest" (the "donor" record). The characteristics of each recipient record were compared to those of each donor record in a pool of donors. When a characteristic between the recipient and a donor were the same, the weight (value) of that characteristic was added to a "score" for that donor. In the end, the donor with the highest score was deemed to be the closest, and was therefore chosen to fill the missing value(s) in the recipient record. If there was more than one donor record with the highest score, one was randomly selected. The pool of donor records was made up in such a way that the imputed value assigned to the recipient record, in conjunction with other non-imputed items from the recipient, would still pass the edits.

Donor imputation was done in three steps. First, household income (G10C) was imputed. This is partly because household income is an important factor in the donor score when imputing key items. Second, the five parent savings variables (Q32A, Q33A, Q34A, Q35A and Q40A) and their corresponding lead-ins (Q20 and Q37A) were imputed. These variables were imputed simultaneously for consistency and coherence. Finally, the other savings by others variable (Q43A) and its corresponding lead-in (Q42) were simultaneously imputed.

The table below shows the rates of imputation for household income and the six items.

Item	All records			Excluding valid skip		
	Imputed	Total	Rate (%)	Imputed	Total	Rate (%)
G10C	1,525	11,544	13	1,525	11,544	13
Q32A	3,025	20,353	15	2,976	8,496	35
Q33A	3,388	20,353	17	3,339	8,496	39
Q34A	4,062	20,353	20	4,013	8,496	47
Q35A	2,367	20,353	12	2,318	8,496	27
Q40A	1,144	20,353	6	749	3,202	23
Q43A	3,398	20,353	17	1,648	2,565	64

Note that the variable G10C is known at the household level while the other savings variables are known at the child level. The rates are calculated based on all records and on records where the value is not valid skip.

The SAEP imputation process worked well. Overall, the process helped fill incomplete responses using the experience of other respondents with similar or identical characteristics, and adds to the number of units available for researchers to analyse.

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 15.6% of children in Canada have a Registered Education Savings Plan (RESP)[5] and this estimate is found to have a standard error of 0.0048. Then the coefficient of variation of the estimate is calculated as:

$$\left(\frac{0.0048}{0.156}\right) X\ 100\% = 3.1\%$$

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

[5] Money saved in RESPs for the child by members of the child's household.

9.0 Guidelines for Tabulation, Analysis and Release

This chapter of the documentation outlines the guidelines to be adhered to by users tabulating, analysing, publishing or otherwise releasing any data derived from the survey microdata 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 Survey of Approaches to Educational Planning (SAEP) was not self-weighting. When producing simple estimates, including the production of ordinary statistical tables, users must apply the proper sampling weight.

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

The SAEP public use microdata file has been set up so that the child is the unit of analysis. The weight on each record is a "child" weight. Estimates of parents, families or households cannot be calculated from the SAEP microdata file.

Before discussing how the SAEP data can be tabulated and analysed, 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 SAEP.

9.3.1 Categorical Estimates

Categorical estimates are estimates of the number, or percentage of the surveyed population possessing certain characteristics or falling into some defined category. The number of children who have savings for post-secondary education or the proportion of children that use Registered Education Savings Plans are examples of such estimates. An estimate of the number of persons possessing a certain characteristic may also be referred to as an estimate of an aggregate.

Examples of Categorical Questions:

Q: Did (child's name)'s parents/guardians set aside a place in the home for him/her to use for studying or doing homework?
R: Yes / No

Q: From the following list, what is your household's highest financial priority?
R: Everyday budget / Savings for post-secondary education / Retirement savings / Other savings

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 amount of educational savings made for children under six years of age. The numerator is an estimate of the total amount of savings for children under six years, and its denominator is the number of children under six years for whom savings were reported.

Examples of Quantitative Questions:

Q: Assuming that (child's name) lives at home, how much do his/her parents/guardians expect that the total cost of his/her education would be?
R: $ |_|_|_|,|_|_|_|

Q: How much money was saved for (child name)'s post-secondary education in 1998?
R: $ |_|_|_|,|_|_|_|

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 $\left(\hat{X}\right)$,

b) summing the final weights of records having the characteristic of interest for the denominator $\left(\hat{Y}\right)$, then

c) dividing estimate a) by estimate b) $\left(\hat{X} / \hat{Y}\right)$.

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 amount of money saved for the post-secondary education of children whose parents hope they will attend university, multiply the value reported in question Q35A (total money saved) by the final weight for the record, then sum this value over all records with Q3 = 4 (university).

To obtain a weighted average of the form $\hat{X} / \hat{Y}$, the numerator $\left(\hat{X}\right)$ is calculated as for a quantitative estimate and the denominator $\left(\hat{Y}\right)$ is calculated as for a categorical estimate. For example, to estimate the average amount of money saved for the post-secondary education of children whose parents hope they will attend university,

a) estimate the total amount of money $\left(\hat{X}\right)$ as described above,

b) estimate the number of people $\left(\hat{Y}\right)$ in this category by summing the final weights of all records with Q3 = 4, then

c) divide estimate a) by estimate b) $\left(\hat{X} / \hat{Y}\right)$.

9.4 Guidelines for Statistical Analysis

The SAEP 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 differ from that which is appropriate in a sample survey framework, with the result that while in many cases the estimates produced by the packages are correct, the variances that are calculated are poor. Approximate variances for simple estimates such as totals, proportions and ratios (for qualitative variables) can be derived using the accompanying Approximate Sampling Variability Tables.

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

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

1) select all respondents from the file who reported SEX = male;

2) calculate the AVERAGE weight for these records by summing the original child weights from the microdata file for these records and then dividing by the number of children where SEX = male;

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

4) perform the analysis for these children using the RESCALED weight

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

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

9.5 Coefficient of Variation Release Guidelines

Before releasing and/or publishing any estimate from the SAEP, 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 weighted rounded estimates.

All estimates can be considered releasable. However, those of marginal or unacceptable quality level must be accompanied by a warning to caution subsequent users.

Quality Level Guidelines

Quality Level of Estimate	**Guidelines**
1) Acceptable	Estimates have a sample size of 30 or more, and low coefficients of variation in the range of 0.0% to 16.5%. No warning is required.
2) Marginal	Estimates have a sample size of 30 or more, and high coefficients of variation in the range of 16.6% to 33.3%. Estimates should be flagged with the letter M (or some similar identifier). They should be accompanied by a warning to caution subsequent users about the high levels of error, associated with the estimates.
3) Unacceptable	Estimates have a sample size of less than 30, or very high coefficients of variation in excess of 33.3%. Statistics Canada recommends not to release estimates of unacceptable quality. However, if the user chooses to do so then estimates should be flagged with the letter U (or some similar identifier) and the following warning should accompany the estimates: "Please be warned that these estimates [flagged with the letter U] do not meet Statistics Canada's quality standards. Conclusions based on these data will be unreliable, and most likely invalid."

9.6 Release Cut-off's for the Survey of Approaches to Educational Planning

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 explains the correct procedure to be used for ratios.

Province and Region	**Acceptable CV 0.0% to 16.5%**	**Marginal CV 16.6% to 33.3%**	**Unacceptable CV > 33.3%**
Newfoundland and Labrador	11,000 & over	3,000 to < 11,000	under 3,000
Prince Edward Island	5,000 & over	1,500 to < 5,000	under 1,500
Nova Scotia	15,500 & over	4,000 to < 15,500	under 4,000
New Brunswick	10,000 & over	2,500 to < 10,000	under 2,500
Quebec	50,000 & over	12,500 to < 50,000	under 12,500
Ontario	49,500 & over	12,500 to < 49,500	under 12,500
Manitoba	15,500 & over	4,000 to < 15,500	under 4,000
Saskatchewan	13,000 & over	3,500 to < 13,000	under 3,500
Alberta	33,000 & over	8,500 to < 33,000	under 8,500
British Columbia	42,000 & over	10,500 to < 42,000	under 10,500
Atlantic Provinces	12,000 & over	3,000 to < 12,000	under 3,000
Prairie Provinces	25,000 & over	6,500 to < 25,000	under 6,500
Canada	**40,500 & over**	**10,000 to < 40,500**	**under 10,000**

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 for children aged 0 to 18 years, which were used to produce the Approximate Sampling Variability Tables for the Survey of Approaches to Educational Planning (SAEP).

Province and Region	Design Effect	Sample Size	Population
Newfoundland and Labrador	2.14	828	129,085
Prince Edward Island	2.30	526	35,859
Nova Scotia	2.42	1,197	221,251
New Brunswick	1.89	1,147	175,739
Quebec	3.26	3,942	1,695,246
Ontario	2.96	6,161	2,863,607
Manitoba	2.39	1,477	279,577
Saskatchewan	2.29	1,638	266,933
Alberta	2.09	1,740	786,608
British Columbia	2.17	1,697	935,429
Atlantic Provinces	2.22	3,698	561,934
Prairie Provinces	2.55	4,855	1,333,118
Canada	**3.07**	**20,353**	**7,389,334**

All coefficients of variation in the Approximate Sampling Variability Tables are approximate and, therefore, unofficial. Estimates of actual variance for specific variables may be obtained from Statistics Canada on a cost-recovery basis. Since the approximate CV is conservative, the use of actual variance estimates may cause the estimate to be switched from one quality level to another. For instance a *marginal* estimate could become *acceptable* based on the exact CV calculation.

Remember: If the number of observations on which an estimate is based is less than 30, the weighted estimate is most likely unacceptable and Statistics Canada recommends not 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 Children 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 Children 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 who have savings for post-secondary education is more reliable than the estimated number of children who have savings for post-secondary education. (Note that in the tables the coefficients of variation decline in value reading from left to right)

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

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

Rule 3: Estimates of Differences Between Aggregates or Percentages

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

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

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

Rule 4: Estimates of Ratios

In the case where the numerator is a subset of the denominator, the ratio should be converted to a percentage and Rule 2 applied. This would apply, for example, to the case where the denominator is the number of children and the numerator is the number of children who have savings for post-secondary education.

In the case where the numerator is not a subset of the denominator, as for example, the ratio of the number of children with savings whose parents hoped they would attend university as compared to the number of children with savings whose parents hoped they would attend college, 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 SAEP are included to assist users in applying the foregoing rules.

Example 1: Estimates of Numbers of Children Possessing a Characteristic (Aggregates)

Suppose that a user estimates that 3,008,755 children have money saved for their post-secondary education by members of their household. How does the user determine the coefficient of variation of this estimate?

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

2) The estimated aggregate (3,008,755) 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 3,000,000.

3) The coefficient of variation for an estimated aggregate is found by referring to the first non-asterisk entry on that row, namely, 1.4%.

4) So the approximate coefficient of variation of the estimate is 1.4%. The finding that 3,008,755 (to be rounded according to the rounding guidelines in Section 9.1) children have money saved for their post-secondary education by members of their household is publishable with no qualifications.

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

Suppose that the user estimates that the parents/guardians of 2,396,350 / 3,008,755 = 79.6% of the children who have money saved for their post-secondary education by members of their household, hope they will go to university. How does the user determine the coefficient of variation of this estimate?

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

2) Because the estimate is a percentage which is based on a subset of the total population (i.e., children who have post-secondary education savings), it is necessary to use both the percentage (79.6%) and the numerator portion of the percentage (2,396,350) in determining the coefficient of variation.

3) The numerator, 2,396,350, does not appear in the left-hand column (the "Numerator of Percentage" column) so it is necessary to use the figure closest to it, namely 2,000,000. Similarly, the percentage estimate does not appear as any of the column headings, so it is necessary to use the percentage closest to it, 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) So the approximate coefficient of variation of the estimate is 1.3%. The finding that the parents/guardians of 79.6% of the children who have money saved for their post-secondary education by members of their household, hope they will go to university can be published with no qualifications.

Example 3: Estimates of Differences Between Aggregates or Percentages

Suppose that a user estimates that the parents/guardians of 1,224,140 / 1,493,156 = 82.0% of girls who have money saved for their post-secondary education by members of their household, hope they will go to university, while the parents/guardians of 1,172,211 / 1,515,599 = 77.3% of boys who have money saved for their post-secondary education by members of their household, hope they will go to university. How does the user determine the coefficient of variation of the difference between these two estimates?

1) Using the Child coefficient of variation table for CANADA in the same manner as described in Example 2 gives the CV of the estimate for girls as 1.1%, and the CV of the estimate for boys as 1.8%.

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

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

where $\hat{X}_1$ is estimate 1 (girls), $\hat{X}_2$ is estimate 2 (boys), 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 $\hat{d}$ = 0.820 – 0.773 = 0.047 is:

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{\left[(0.820)(0.011)\right]^2 + \left[(0.773)(0.018)\right]^2} \\ &= \sqrt{(0.000081) + (0.000194)} \\ &= 0.017\end{aligned}$$

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d}$ = 0.017 / 0.047 = 0.362.

4) So the approximate coefficient of variation of the difference between the estimates is 36.2%. This estimate is considered unacceptable and Statistics Canada recommends this estimate not be released. However, should the user choose to do so, the estimate should be flagged with the letter U (or some similar identifier) and be accompanied by a warning to caution subsequent users about the high levels of error, associated with the estimate.

Example 4: Estimates of Ratios

Suppose that the user estimates that the parents/guardians of 1,224,140 girls who have money saved for their post-secondary education by members of their household, hope they will go to university, while the parents/guardians of 1,172,211 boys who have money saved for their post-secondary education by members of their household, hope they will go to university. The user is interested in comparing the estimate of girls versus that of boys 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 girls with educational savings whose parents/guardians hope they will go to university. The denominator of the estimate ($\hat{X}_2$) is the number of boys with educational savings whose parents/guardians hope they will go to university.

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

3) The numerator of this ratio estimate is 1,224,140. 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, 3.1%.

4) The denominator of this ratio estimate is 1,172,211. 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, 3.1%.

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:

$$\begin{aligned}\alpha_{\hat{R}} &= \sqrt{(0.031)^2 + (0.031)^2} \\ &= \sqrt{0.000961 + 0.000961} \\ &= 0.044\end{aligned}$$

6) The obtained ratio of girls versus boys who have money saved for their post-secondary education by members of their household and whose parents/guardians hope they will go to university is 1,224,140 / 1,172,211 which is 1.04:1 (to be rounded according to the rounding guidelines in Section 9.1). The coefficient of variation of this estimate is 4.4%, which is releasable with no qualifications.

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 differences 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 $\left(CI_{\hat{x}}\right)$:
>
> $$CI_{\hat{x}} = \left(\hat{X} - t\hat{X}\alpha_{\hat{x}}, \hat{X} + t\hat{X}\alpha_{\hat{x}}\right)$$
>
> 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 who have money saved for their post-secondary education by members of their household and whose parents/guardians hope they will go to university (from Example 2, Section 10.1.1) would be calculated as follows:

$\hat{X}$ = 79.6% (or expressed as a proportion 0.796)

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.

$CI_{\hat{x}}$ = {0.796 - (2) (0.796) (0.013), 0.796 + (2) (0.796) (0.013)}

$CI_{\hat{x}}$ = {0.796 – 0.021, 0.796 + 0.021}

$CI_{\hat{x}}$ = {0.775, 0.817}

With 95% confidence it can be said that between 77.5% and 81.7% of children who have money saved for their post-secondary education by members of their household, parents/guardians hope that they will go to university.

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 girls who have money saved for their post-secondary education by members of their household, whose parents/guardians hope they will go to university, and the proportion of boys who have money saved for their post-secondary education by members of their household, whose parents/guardians hope they will go to university. From Example 3, Section 10.1.1, the standard error of the difference between these two estimates was found to be 0.017. Hence,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0.820 - 0.773}{0.017} = \frac{0.047}{0.017} = 2.76$$

Since t = 2.76 is greater than 2, it must be concluded that there is a 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 Survey of Approaches to Educational Planning 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 money saved in Registered Education Savings Plans would be greater than the coefficient of variation of the corresponding proportion of children with savings in Registered Education Savings Plans. Hence if the coefficient of variation of the proportion is unacceptable (making the proportion not releasable), then the coefficient of variation of the corresponding quantitative estimate will also be unacceptable (making the quantitative estimate not releasable).

Coefficients of variation of such estimates can be derived as required for a specific estimate using a technique known as pseudo replication. This involves dividing the records on the microdata files into subgroups (or replicates) and determining the variation in the estimate from replicate to replicate. Users wishing to derive coefficients of variation for quantitative estimates may contact Statistics Canada for advice on the allocation of records to appropriate replicates and the formulae to be used in these calculations.

10.5 Coefficient of Variation Tables

See SAEP1999_Master_Child_CVTabsE.pdf for the coefficient of variation tables to be used with the child public use microdata file (PUMF) and the child master file.

11.0 Weighting

Since the Survey of Approaches to Educational Planning (SAEP) used a sub-sample of the Labour Force Survey (LFS) sample, the derivation of weights for the survey records is clearly tied to the weighting procedure used for the LFS. The LFS weighting procedure is briefly described below.

11.1 Weighting Procedures for the Labour Force Survey

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

Basic Weight

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

Cluster Sub-weight

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

Stabilization Weight

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

Non-response

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

In other cases, non-response is compensated for by proportionally increasing the weights of responding households. The weight of each responding record is increased by the ratio of the number of households that should have been interviewed, divided by the number that were actually interviewed. This adjustment is done separately for non-response areas, which are

defined by employment insurance economic region, type of area, and rotation group. It is based on the assumption that the households that have been interviewed represent the characteristics of those that should have been interviewed within a non-response area.

Labour Force Survey Sub-weight

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

Sub-provincial and Province-Age-Sex Adjustments

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

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

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

11.2 Weighting Procedures for the Survey of Approaches to Educational Planning

The principles behind the calculation of the weights for the Survey of Approaches to Educational Planning are identical to those for the LFS. However, further adjustments are made to the LFS sub-weights in order to derive a final weight for the individual records on the SAEP microdata file.

1) An adjustment to account for the use of a five-sixth sub-sample, instead of the full LFS sample.

2) An adjustment to account for the additional household non-response to the supplementary survey, i.e., non-response to the SAEP for households who did respond to the LFS or for which previous month's LFS data was brought forward. This procedure is similar to the LFS non-response weight adjustment, but grouping are based on different variables.

3) A readjustment to account for independent projections of the number of households by province and household size (one, two and three or more persons), after the above adjustments are made.

Child weights were derived as follows:

4) An adjustment to the non-response adjusted weight from adjustment 2) above, to account for the sub-sampling of children in households with more than three children.

5) A readjustment to account for independent projections of the number of persons by province, age group and sex. The age groups used were 0 to 5, 6 to 12, 13 to 15, and 16 to 18.

Adjustments 1) and 2) are taken into account by multiplying the LFS sub-weight for each SAEP record by:

$$= \left(\frac{6}{5}\right) * \left(\frac{\sum LFS\ sub-weights\ \ from\ each\ household\ \ in\ rotation\ \ groups\ 1,2,3,5\ and\ 6\ responding\ \ to\ LFS}{\sum LFS\ sub-weights\ \ from\ each\ household\ \ responding\ \ to\ SAEP}\right)$$

to obtain a non-response adjusted household sub-weight (WEIGHT1).

Independent estimates are available monthly for the number of households containing one, two and three or more persons by province. These are population projections based on the 1996 Census data, records of births and deaths, and estimates of migration. A calibration was performed to arrive at the final household weight, ensuring that the final weights it produces sum to the census projections for these auxiliary variables. This improves the reliability of estimates that can be produced by the Survey of Approaches to Educational Planning. Adjustment 3) was calculated by multiplying WEIGHT1 for each SAEP respondent by:

$$= \frac{projected\ number\ of\ households\ within\ province\ and\ household\ size\ group}{\sum WEIGHT1\ for\ SAEP\ respondent\ households\ within\ province\ and\ household\ size\ group}$$

The resulting weight is the final household weight that appears on the Survey of Approaches to Educational Planning household master file held by Statistics Canada. Note that the household weight is not available on the public use microdata file.

Adjustments 4) and 5) are performed in order to calculate a weight for each child for whom there is a SAEP response. Adjustment 4) is performed by multiplying WEIGHT1 by:

$$= \frac{number\ of\ children\ 18\ and\ under\ in\ the\ household}{\min imum\ number\ of\ children\ 18\ and\ under\ in\ the\ household\ \ up\ to\ a\ \max imum\ of\ 3}$$

to obtain a sub-sampling adjusted child weight WEIGHT3.

Independent estimates are also available monthly by age, sex and province. A calibration was performed to arrive at the final child weight, ensuring that the final weights it produces sum to the census projections by province, sex and age groups. The age groups were 0 to 5, 6 to 12, 13 to 15, and 16 to 18.

This improves the reliability of estimates that can be produced by the Survey of Approaches to Educational Planning. Adjustment 5) was calculated by multiplying WEIGHT1 for each SAEP respondent by:

$$= \frac{projected\ \ number\ \ of\ \ children\ \ within\ \ province\ ,sex\ and\ \ age\ group}{\sum WEIGHT\ 3\ for\ the\ SA\ EP\ respondent\ s'\ \ children\ w\ ithin\ \ province,\ sex\ and\ \ age\ group}$$

The resulting weight (FINWT) is the final child weight that appears on the Survey of Approaches to Educational Planning child master file and public use microdata file.

Note that the variable FINWT (12,4) on the child master file and child public use microdata file has a physical decimal four bytes from the right (9999999.9999).

12.0 Questionnaires

12.1 The Labour Force Survey Questionnaire

The Labour Force Survey questionnaire (LFS_QuestE.pdf) is used to collect information on the current and most recent labour market activity of all household members 15 years of age or older. It includes questions on hours of work, job tenure, type of work, reason for hours lost or absent, job search undertaken, availability for work, and school attendance.

12.2 The Survey of Approaches to Educational Planning Questionnaire

The Survey of Approaches to Educational Planning (SAEP) questionnaire was used in October 1999 to collect the information for the supplementary survey. The file SAEP1999_QuestE.pdf contains the English questionnaire.

The 1999 Survey of Approaches to Educational Planning contains a total of 171 questions organized into seven sections (A to G). The first section (A) has a screening question to ensure that respondents will answer only the questions that are relevant to their situation. The next three sections (B, C and D) ask about the school experiences and educational savings for three children in the household. Section E asks about educational savings for any other children 18 years of age or younger in the household. Section F asks questions about educational savings for children who do not live in the household. The final section (G) asks questions of a more general nature to profile certain household characteristics.

Section A: Screening questions

The questions in this section are designed to determine whether any children 18 years of age or younger usually live in the household. Based on the answers to these questions, the respondent is directed to the appropriate section of the questionnaire. In the case where there are more than three children 18 years of age or younger in the household, three children are randomly selected.

Section B: Child 1

This section asks questions about the first child in the household who is 18 years of age or younger. Questions in this section ask about the child's school experiences in the past school year, how much the parents/guardians expect the child's post-secondary to cost and how they expect to pay for it. If savings have been put aside for the child, questions are asked about amounts saved, how the savings are invested and how much of the savings are in Registered Education Savings Plans.

Section C: Child 2

This section asks the same questions as Section B, for the second child 18 years of age or younger in the household.

Section D: Child 3

This section asks the same questions as Section B, for the third child 18 years of age or younger in the household.

Section E: Remaining children in the household: Savings for post-secondary education

This section asks about educational savings for any other children 18 years of age or younger who live in the household. Questions collect information about the amount and type of savings.

Section F: Children outside the household

This section asks about educational savings for children who do not live in the household. Questions collect information about the amount and type of savings.

Section G: General information

The questions in this section collect information about resources available in the household, ethnicity, language, financial priorities and income.

13.0 Record Layouts with Univariate Frequencies

See SAEP1999_PUMF_Child_CdBk.pdf for the record layout with univariate counts for the child public use microdata file (PUMF).

See SAEP1999_Master_Child_CdBk.pdf for the record layout with univariate counts for the child master file.

13.1 Unit of Analysis

The Survey of Approaches to Educational Planning (SAEP) file has been set up so that the child is the unit of analysis. The file contains one record per child. The weight that can be found on each record is a "child" weight. When weighted, the records in this file will sum up to the total number of children 0 to 18 years old in Canada. Estimates of parents, families or households cannot be made from the SAEP microdata file.

As the microdata file is at the child-level, it excludes households that did not have any children living with them at the time of the survey. The microdata also does not include variables from Section E (savings for the remaining children in the household when there are more than three children in the household) or from Section F (savings for children who live outside the respondent's household). Some derived variables on the microdata file are created using data from these sections, namely, derived variables relating to household savings. More detailed information about households without children and Sections E and F can be obtained by custom tabulations available through Special Surveys Division's Client Services unit (see Chapter 1.0).

13.2 Availability of Parent Information

Socio-demographic and labour force information about a child's mother and/or father is available for most children. However, there are two circumstances in which data is not available for one or both parents. First, if a parent does not live in the same household as the child, data is not available for that parent. Second, if a child-parent relationship cannot be derived from the Labour Force Survey data, data is not available for the parents.

The parent-child relationship is derived from information collected by the Labour Force Survey about the relationship of people to each other in an economic family. An economic family is a group of two or more people who live in the same dwelling and are related to each other by blood, marriage, common-law or adoption. The relationship of economic family members to each other is based on the Labour Force Survey respondent, or "reference person". This means that we know the relationship of the reference person to everyone in the economic family, but we do not know the relationship of every family member to every other family member. For example, take the case of a child who lives with his mother and his aunt. If the mother is the reference person, her relationship to the child is recorded in the Labour Force Survey as "son/daughter". However, if the aunt is the reference person, her relationship to the child is recorded as "brother or sister or other relative" and her relationship to the mother is recorded as "brother or sister or other relative." The relationship between the mother and child is unknown. In the second case, there will be no information about the child's parents on the public use microdata file because a parent-child relationship could not be determined.

Another variable that is derived from the viewpoint of the Labour Force Survey respondent is the variable EFAMTYPE which describes the economic family type. The economic family type is calculated from the viewpoint of the Labour Force Survey reference person, not from the viewpoint of the child. If a child lives with her mother and father and her father is the LFS reference person, the economic family type would be "husband-wife". However, if the child lives with her mother and her grandparents and her grandmother is the LFS reference person, the economic family type would be "husband-wife". However, in this case the "husband-wife" designation refers to the child's grandparents, not to the child's parents.

13.3 Naming Convention for Variables

Sections B, C and D in the Survey of Approaches to Educational Planning questionnaire collected information about the children in the household. On the questionnaire, the numbering was based on the section. For example, Section B collected information about the first child in the household and the first question was numbered B1.

On the microdata file, there is one record per child. The variable names for Sections B, C and D have been changed to Q1 (equivalent to B1, C1 or D1), Q2 (equivalent to B2, C2 or D2), etc.