Centre for Education Statistics, Canada

Centre des statistiques sur l'éducation, Canada

School Leavers Follow-up Survey - 1995

Microdata User's Guide

May 1997

Statistics Canada Statistique Canada 81M0016XDB

Canada

School Leavers Follow-up Survey

Microdata User's Guide

TABLE OF CONTENTS

1. Introduction

In the late 1980s, Human Resources Development Canada (HRDC), formerly Employment and Immigration Canada, commissioned Statistics Canada to conduct a survey to identify the characteristics of high school leavers and the consequences of leaving high school. Between April and June 1991, computer assisted telephone interviews (CATI) obtained demographic and background information on school experiences and post-school labour market experiences from 9,460 young Canadians aged 18 to 20. The two primary objectives of the School Leavers Survey (SLS) were to: 1. develop comparative profiles of secondary school students who successfully completed high school (graduates), who were still attending (continuers) and who left school before graduating (leavers) and 2. to establish high school leaving rates.

In early 1994, HRDC requested that Statistics Canada re-interview the 1991 survey respondents. The School Leavers Follow-up Survey (SLFS) was conducted between September and December 1995, again using the CATI instrument. The primary objective of the Follow-up was to study the school-work transitions of young people beyond high school, particularly: 1. from the end of high school to the first job and 2. after the first job (i.e. various school-work and work-work transitions).

This document describes the methodology used by Statistics Canada in conducting the School Leavers Survey and School Leavers Follow-up Survey. It is also intended to be used as a microdata user's guide to help facilitate the analysis of the data. Please address any inquiries about the data set or its use to the **Centre for Education Statistics, Canada** at (613) 951-7474 or by facsimile at (613) 951-9040.

2. Background

2.1 School Leavers Survey, 1991

In light of traditionally high estimates of secondary school non-completion (30%+), HRDC commissioned Statistics Canada to conduct a School Leavers Survey to estimate the magnitude of the problem and to identify the circumstances associated with dropping out.

Using the Family Allowance files as the sampling frame, a stratified random sample of 18,000 18 to 20-year-olds was selected. A computer assisted telephone interview (CATI) obtained demographic/background, school experience, and post-school labour market and other outcome measures, along with whether respondents were in school (continuers), had graduated (graduates), or had left before graduating (leavers). A total of 10,782 individuals were traced and contacted (60% contact rate), and of those, 9,460 were interviewed (88% of those traced). The respondents represent 184,000 leavers, 241,000 continuers and 711,000 graduates.

Leaving School (Gilbert et al, 1993) and ongoing multivariate analyses of the SLS have generated new findings which provide important policy data:

- the high school leaver rate is in the 18-20% range.
- the rate for men is higher than for women.
- rates are higher in the eastern provinces and lower in the West
- the gap between males and females is greater in the East and less pronounced in the West.
- leavers were more likely than graduates to be unemployed.
- more male leavers than graduates were employed in blue collar primary industries and processing occupations, and fewer were in white collar clerical, sales and service jobs.
- female leavers and graduates were clustered in clerical, sales and service jobs, with female leavers tending to be in service jobs and female graduates tending to be in clerical and related positions.
- many male leavers and graduates worked 50 or more hours per week.
- leavers were more restricted in their job search activities and anticipated less help than graduates. Leavers had difficulty filling out job applications and their reading, writing and mathematics skills limited their job opportunities. In particular, female leavers said that their basic mathematics skills restricted their employment chances.
- leavers relied more upon unemployment insurance and social assistance than graduates.
- both leavers and graduates, particularly females, were not doing well financially. Leavers were generally dissatisfied with their financial situation.

2.2 School Leavers Follow-up Survey, 1995

Young adults today can anticipate a number of major education and work transitions over the course of their life. The traditional boundaries between work and school have become blurred, with many students working while studying. Many of today's youth will also have 4 or 5 distinct careers and many jobs within each career area during their active labour force period. At the same time the economy is being transformed from unskilled, manual, blue collar manufacturing jobs to white collar, knowledge and service occupations. Economic restructuring means that a new

range of skills, competencies and performances will likely be required in future labour markets.

The first key transition for today's youth is life beyond high school. Up to that point in time, young Canadians have been in day-care, pre-school, kindergarten or school for 11-14 years, or about 70% of their early formative lives. The departure from high school represents a first major break with communities and associations of the past (families, friends and schooling) and an emerging connection to communities and associations of the future (employment, higher education and new economic and social networks). What happens upon leaving high school is a milestone event in the life cycle, with important long-term consequences. How youth cope with the transition from high school, and the early jobs, occupations and careers they have, including work skills and competencies, may influence future economic success or failure.

The 1991 School Leavers Survey began to examine the transition from school to work, entry points into the labour force and associated factors. Given that the sample frame was 18-20 year olds, there was only a limited opportunity to evaluate the impact of secondary school graduation upon entry-level jobs. The **Leaving School** (Gilbert et al, 1993) analyses do not show great differences in employment earnings between leavers and high school graduates. However, with the passage of time and the acquisition of skills and experience, greater earnings differences were anticipated.

When the SLS was conducted in 1991, 47% of the male 18- year- old and 43% of the female 18- year- old respondents were still attending school. In 1995 most of these 1991 continuers have become graduates or leavers (with a small percentage still attending high school). The Follow-up in 1995 provides up-dated information on the education and employment status of high school graduates, leavers and continuers.

Early findings from the SLFS in **After High School** (Frank, 1996) have provided the initial results:

- there are more graduates and fewer leavers than in 1991.
- high school graduation is the gateway to further education.
- unemployment rates are highest among high school leavers.
- most young people expect to take further education or training.

3. Objectives

3.1 School Leavers Survey, 1991

The primary purpose of the survey was to establish a high school leaver rate for Canada and the provinces, as well as develop comparative profiles of three groups of secondary school attendees: those who successfully completed high school, those still in high school and those who left high school before receiving a diploma or certificate. It was to obtain information on factors affecting the potential for a successful transition to the labour market, labour market experience and quality of life.

The survey was designed to address questions of what environmental, social, attitudinal, personal and economic factors contribute to the early departure of some students from school to the labour force, their early labour market experiences and their quality of life. As it was seen at the outset, results from a Leavers Survey would allow analysts to better understand the relationship between leavers and their backgrounds and subsequent labour market experiences. It was hoped that the results would provide school counsellors with better information on the impact of leaving high school. With the information collected by the survey, it would be possible to have a better understanding of the factors behind an individual's behaviour, career choices and successes or failures during the period prior to their entry into the labour market through a set of current and retrospective questions.

More specifically, questions in the survey instrument were designed to address the following types of questions and included:

<u>Transitions</u>

- In what grade do students tend to withdraw from educational institutions? At what rate?
- What reasons do they give for leaving the school system?
- Are there any critical factors related to early leaving?
- What are the transition patterns from school to the labour market?
- What types of training have they experienced since leaving the formal education system?
- What are the barriers to further training?
- What is their current attitude to leaving; i.e. in retrospect, given their experience since leaving school, would leavers now rather have completed their education?

<u>Labour Market</u>

- What occupations do early leavers (and others) obtain?
- To what extent do early leavers act as a feeder group to further training such as apprenticeship groups?
- How does the social and economic environment influence their entry into the labour market?
- After leaving the formal school system, what additional training do individuals get and what impact does this have on their success/failure in the labour market?
- What are the methods of job search used by school leavers? and how effective are they?
- How much help do these individuals need and/or receive in the process of job search?
- What are the individuals expectations for their future in the labour market?

Quality of Life

- To what extent does their lack of specific skills such as reading, numeracy, or computer handling act as a road block to advancement in the labour force?
- What are their current living arrangements? (own home, parent or relatives home, shared accommodation with friends etc...)
- What is their current financial situation, income (sources), expenditures etc.?
- What is their image/satisfaction with their life style?

3.2 School Leavers Follow-up Survey, 1995

The primary objective of the Follow-up was to study the school-work transitions of young people beyond high school, particularly: 1. from the end of high school to the first job and 2. after the first job (i.e. various school-work and work-work transitions). The additional information will permit better estimation of leaver rates, and more refined analyses of labour market experience and the transition between high school, the labour force and additional education and training.

The Follow-up provides answers to strategic issues such as:

School Leavers:

- How many of the 1991 school leavers returned to school and obtained their high school certificates?
- How many of the 1991 continuers left school without a diploma?
- What subsequent training, education or learning have leavers received?
- What is the medium-term labour market experience of high school leavers (number of jobs, type of job and sector, skills used, hours, earnings, career progression)?
- What is the extent of UI or social assistance dependency?
- How long did they keep their first job and if they left that job what were the reasons?

High School Graduates:

- What subsequent training, education or learning have high school graduates received?
- What skills, abilities, competencies and work-related values and attitudes have they acquired as a result of the training, education or learning?
- How many of the high school graduates entered post-secondary education or training and did not complete their programs?
- What is the job-entry experience of post-secondary leavers compared to graduates (number of jobs, type of job and sector, skills used, hours, earnings, career progression)?
- How long did they keep their first job and if they left that job what were the reasons?

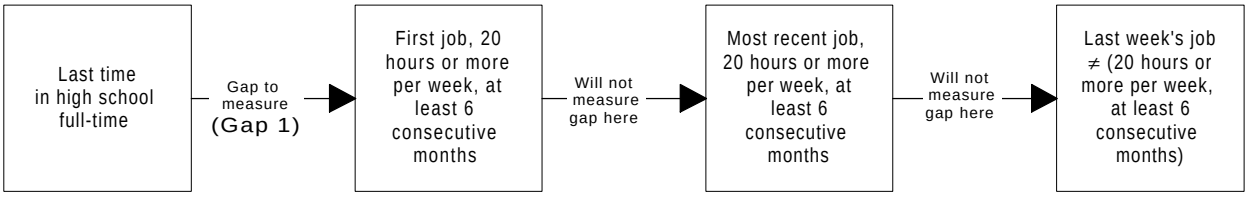
4. Concepts and Definitions

4.1 Doing Labour Market Analysis with the SLFS

The definition of concepts and identification of data to be gathered in the SLFS represented major challenges in developing labour market questions. For example, it was necessary to establish a reference period for the survey. This period begins from the point at which a respondent was last in high school full-time and ends the week preceding the interview. Measuring all jobs held by respondents was not feasible. Therefore, the concept of "reference job" was developed. The definition of "reference job" and other concepts used in the labour market section of the Follow-up Survey are discussed in detail in the following definition section.

Figure 1 shows the labour market pattern that was measured. The reference period began when the full-time student exited high school and ended in the week preceding the interview. The length of the reference period will vary among respondents.

Figure 1 Labour Market Pattern



Clearly, gathering detailed information on each job held by respondents since the end of high school was not possible. Accurate recall of information would be too much to expect of those who held numerous jobs over a long period of time. It was, therefore, decided to concentrate on a few key jobs (reference jobs) which make it possible to study progressions in the labour market. The criterion for a "reference job" was its duration – expressed as the number of hours worked per week and the number of months worked. A reference job is one of sufficient duration, that is, a job involving at least 20 hours of work per week for a period of at least six consecutive months.

More workers hold full-time than part-time jobs. However, in recent years, a growing number of people have been working part-time (by choice or out of necessity). For purposes of the survey, it was decided not to target full-time work only (work of 30 hours or more per week), but to obtain information about part-time jobs involving a sufficient number of hours of work, that is, at least 20 hours per week.

Short-term jobs are sometimes defined as lasting about six months or less. It was decided that jobs targeted for purposes of the survey would have been held for a period of six consecutive months or more.

Detailed information was gathered on a maximum of three separate jobs:

- the first reference job after leaving high school full-time;
- the most recent reference job;
- the job held the week before the interview.

Users of the survey data will likely require information on respondents' labour force status at the time of the interview. The Follow-up survey determined whether respondents were employed, unemployed or not in the labour force the week prior to being interviewed. Information on the job held that week was also gathered, even if the job did not meet the reference job criterion.

If the job held the week prior to the interview did not meet the reference job criterion, information was gathered on the most recent reference job held by the respondent.

To study the initial transition from school-to-work, data on the first reference job held after last attending high school full-time was collected.

For certain respondents, information could not always be obtained on these three jobs. Some had not yet held any job and others held one job only. The various possibilities that were encountered during the Follow-up survey are represented by the five patterns in Figure 2.

The data file for the SLFS has been structured to assist users in performing labour market analysis. Data on the labour market status last week and job held last week are grouped separately in the layout and begin in field 310 and end in field 360. Data on reference jobs are also shown in a separate section beginning in field 361 and ending in field 404. **It is important to note that if the job held last week was a reference job, the data in these sections overlap; that is, last week's job information is repeated where appropriate in the reference job section.**

The reference job section of the layout contains data for a maximum of two reference jobs- the first and most recent. Where data are available for only one reference job, these data are always shown in the first reference job fields (they are not repeated in the most recent reference job fields). Where data are available for two reference jobs, there is information in both the first and most recent reference job sections.

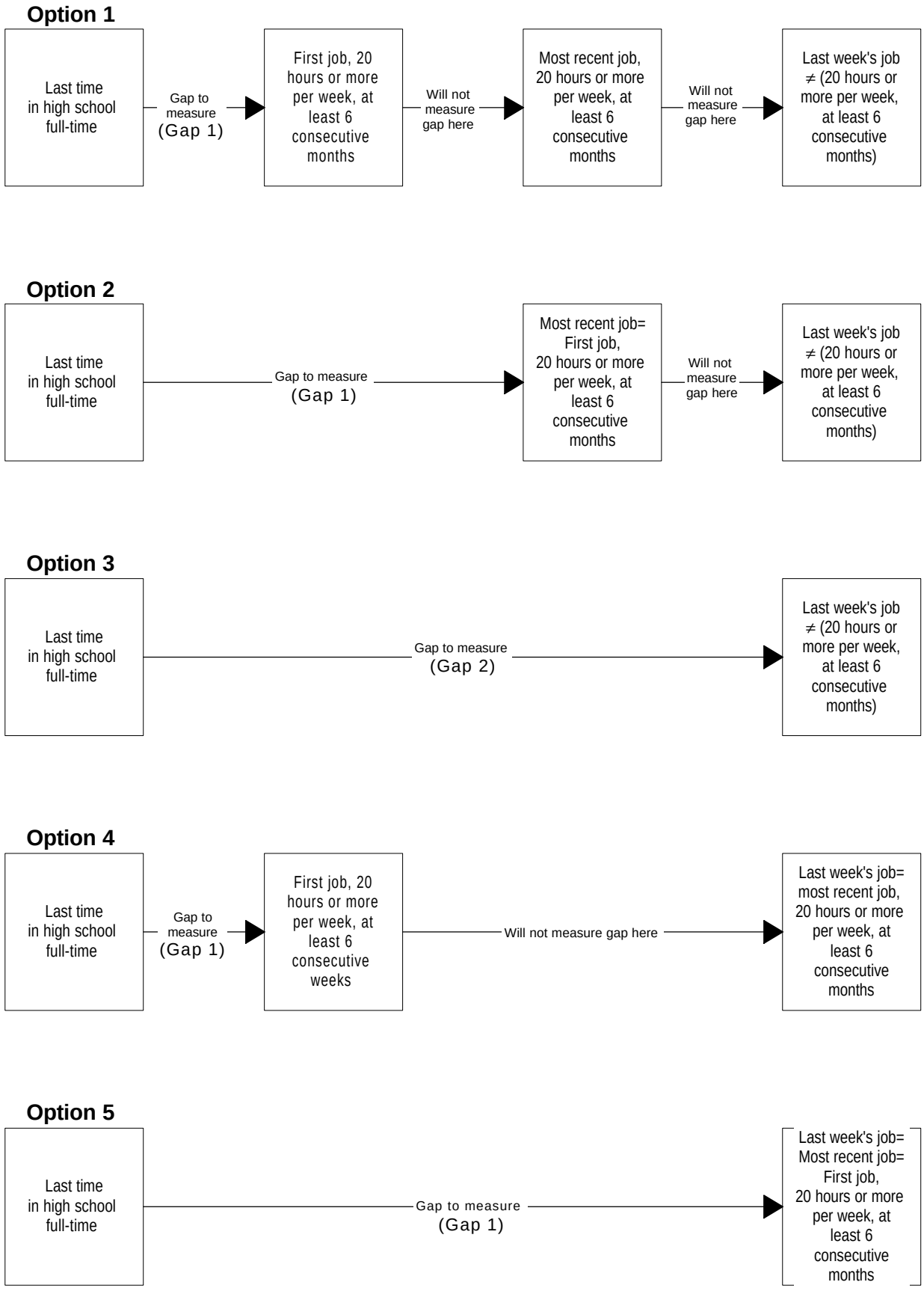
A variable has been created on the file to assist users in determining the number of reference jobs held and the source of the reference job information from the questionnaire. That is, the variable specifies if there are data available for more than one reference job and if the reference job information came from last week's job. This variable, REFSOURC, is found in field 364 and should help users in their analysis of reference jobs.

Period between two events: GAPS

Not only did the Follow-up survey gather data on jobs, it also collected information on the transitions process. Respondents were asked about their labour market and education/training activities during the period between the last time they were high school full-time and their first reference job (see figure 2, gap 1). This period is referred to in the questionnaire and record layout as GAP 1. Data for GAP 1 are found in fields 405 to 418.

Where no reference jobs existed, respondents were asked about their labour market and education/training activities since they were last in high school full-time up to the beginning of the week prior to being interviewed (see figure 2, gap 2). If a respondent was employed (at a non-reference job) during the week prior to being interviewed the period refers to the time between the date last in high school full-time and the beginning of last week's job. Otherwise, the period refers to the time between the date last in high school full-time and the date of the interview. This period is referred to in the questionnaire and record layout as GAP 2. Data for GAP 2 are found in fields 419 to 432.

Figure 2 Five Labour Market Patterns Measured in the SLFS



4.2 Skill Assessment and the SLFS

Given the changing nature of the workplace and the emphasis on human resource development, HRDC and Statistics Canada recognized the importance of providing skill assessment measures on the SLFS. Twenty-four questions that measured **the frequency of skill use** across a variety of dimensions and 6 self-assessment measures were included in the SLFS. The 24 skill use questions asked respondents about the frequency of certain activities related to different basic skills while the 6 self-assessment questions asked respondents to rate their abilities in these basic skills.

Skills Chosen for Evaluation in the SLFS

The following list of skills constitute those generally considered as being sought by employers (see The Conference Board of Canada's Employability Skills Profile, the Corporate-Higher Education Forum's Learning Goals and Jones, 1993 for a discussion of skills). Undoubtedly, the list is not exhaustive and the emphasis placed on each of the various skills will vary over different occupations.

The following list contains a proposed definition for each skill:

Basic Skills (Functionally Literate - Reading, Writing and Numeracy)

At the various levels of the educational structure, students who are graduating are expected to have attained a certain level of proficiency in reading, writing, fundamental numeracy [numeracy = arithmetic (+, -, x, /) and mathematics (algebra, trigonometry, etc.)]. This proficiency would vary, of course, with the level of education achieved.

Communications

Communications is considered as the ability to understand others in all forms of communication (written and spoken), as well as the ability to make others understand the information being delivered. There are four aspects to the concept: giving and receiving written communications and giving and receiving spoken communications.

Learning (Ability to Learn)

What is known today is not likely to be sufficient to maintain a job or position in the future. Employers are looking for people who are able to continue to expand and develop their abilities and skills (life long learning).

Workplace Interpersonal Including Team Work

In today's workplace it is increasingly common place for people to work together, and to work in and with groups. This skill, however, is at best difficult to both establish and then measure.

There are a number of other qualities that employers will look for in the potential employee that would not be considered as skill sets, yet employers would consider them as qualities that should be possessed by the potential employee for many of the positions they fill, such as attitude and behaviour sets, responsibility, and adaptability.

Development of Indicators for the SLFS

There are few well developed tests available to establish levels of attainment, proficiency or even the existence of these skills. For those skills where tests have been developed, such as literacy, the tests are not only cumbersome but expensive to administer. In most surveys, skills are only one of a number of variables that must be collected. Thus, the fully developed and calibrated measures that currently exist for a few skills are not appropriate for use in most surveys.

As a result, there is a need to develop proxy measures for the various sets of skills. Research done to date on the measurement of skills suggests that proxy measures based on behaviour, as opposed to self-evaluation, would be a useful approach to skill assessment. Substantiation of this hypothesis comes from work done in the area of literacy (Neice and Adsett, 1990). Self-assessment questions may not give an accurate measure of the skill itself, but measure another aspect of that skill (see Jones, 1991).

Using the approach of asking respondents about their behaviour (as opposed to asking them to self-estimate their competence; cf. Stan Jones' work on literacy and numeracy) a number of possible skill proxies scales were created for the SLFS. A battery of questions in each area were constructed on the assumption that proxies could be developed by asking respondents if they *use* or do particular things, if they select or *choose* among a variety of alternatives, and if they create new ways of acting/behaving (Jones, 1993).

A disadvantage of the behaviour-based approach was that depending on how the questions were worded, or how the person interpreted the question, or even what the person was currently involved in his or her life, the wrong rating could be established. That is, a person may answer **No** to all of the questions and actually have the skill but simply not be doing (using) it at the time he/she was interviewed. There is therefore, a need to evaluate the responses to these questions in the context of responses to other questions, such as level of education achieved.

The inclusion of self-evaluation questions (asking respondents to rate their abilities) allows for further assessment of potential under-representation of respondents on some skills.

Using the Skill Measures on the SLFS

The SLFS data file contains a summary **skill use score** for each respondent and a measure of the standard error of that score. The scores are Rasch scores and can be treated as interval level data. The scores are centred on 0 so there are negative scores. Users who are bothered by this may transform the scores in any way that preserves the interval structure of the data. A popular transformation is:

NewScore=(Skillsco * 50)+250

This transformation will centre the scores on 250. Because the skill questions were structured so that the 'high skill' answer was scored 1 and the 'low skilled' answer was scored 5, high performance is indicated by low scores (think of it as golf). Users who are uncomfortable with this can simply change the sign of the score (before transformation if that is to be done).

The scale was derived from six groups of four questions; each group was intended to represent a different skill. The 24 questions actually scale well on one scale and this overall skill use score is what appears on the file. The reliability of the overall Rasch scale was found to be .90 (Jones, 1997).

It is unlikely that the scores represent skill ability. In recent tests of the measures, the correlation of these scores with direct tests of skill were relatively low (Jones, 1997). These scores are probably best thought of as 'skill use'

scores. Similarly, the correlation of the 6 self-assessment measures with direct tests of skill were relatively low. For more information on the skill scale contact Statistics Canada at the number indicated on the inside cover of the user's guide.

4.3 Definitions

The following section outlines the concepts and definitions used in both the School Leavers Survey and the School Leavers Follow-up Survey. Copies of both survey questionnaires appear in Appendix A and B of this document. These questionnaires are simply a print-out of the Computer Assisted Telephone Interview (CATI) program that was designed to collect the information from respondents.

Alternate Programs

High school courses designed for students who otherwise might dropout. Students study at their own pace often at non-traditional school sites, such as shopping malls.

Alternative School

A school offering a provincially-approved curriculum that uses different teaching methods or places emphasis on teaching cultural identity. Parents frequently work with the teachers in the classroom and in planning programs.

Apprenticeship Program

Apprenticeship programs lead to journeyman status in several designated trades. Skills and knowledge are provided through on-the-job experience with short periods of formal instruction. Examples of apprenticeship trades are: boiler maker, millwright, electrician, plumber, machinist, heavy duty equipment mechanic, cook, etc.

Commercial School

These are private schools, who receive no public funding and who are licensed by a province, which are engaged in providing professional and vocational training for profit.

Community Colleges

Community Colleges include community colleges, colleges of applied arts and technology (CAATS in Ontario), collèges classiques or CEGEPS in Quebec, technical institutes, hospital and regional schools of nursing, one teachers' college, and establishments providing technological training in specialized fields. Community colleges offer career programs of one to four years. Some also provide one- or two-year academic programs which prepare a student to proceed to university.

1991 Continuers

Respondents who are continuing their studies at elementary, junior high or high school. Respondents continuing with postsecondary are not classified as continuers in this survey, but as graduates.

1995 Continuers

Respondents who have not graduated from high school and are formally registered and currently taking a course or

a program (even if he/she wasn't in classes last week) in high school, junior high or elementary school at the time of the survey (1995). Adult high schools are included but correspondence courses are excluded. Respondents continuing with postsecondary are not classified as continuers in the SLFS, but as graduates.

Correspondence Program

An educational or training activity that does not require students to attend a school, college or university. Mail, radio, television or other media communications deliver the instruction.

Educational Attainment

Refers to the degree, certificate, or diploma received by an individual.

Elementary School

The educational structure varies across the provinces. In general, at the elementary grade level, education is general and basic, and as a minimum includes grades kindergarten through six.

Ever Dropped Out

Respondents who have dropped out of elementary/secondary school at one time or another during their school career. Included are high school graduates who have dropped out of school at some point before they graduated, continuers who have dropped out of school at some point but who have returned, and leavers.

Field of Study

Field of study codes up to and including code 482 are based on 1991 Census coding structure of major field of study.

Full time/part time job

A full-time job refers to the circumstance in which the respondent worked 30 or more hours per week. Working less than 30 hours per week was classified as part-time.

Full-time/part-time studies - High School, Junior High or Elementary School

Students are classified as full-time or part-time based on the number of courses or hours of instruction. If respondents wondered about the meaning of full-time, they were informed that it represented 5 hours or more per day at school. Part-time represented less than 5 hours per day at school.

Full-time Program/Part-time Program - Postsecondary

Full-time, part-time status is determined by the education institution. All schools classify their students as being full-time or part-time depending on the number of courses in which they are enrolled. Hence, whether a person was marked full-time or part-time depends on how he/she was classified by the school, college or university which he/she attended.

Gap 1

The period between the last time in high school full-time and the beginning of the first reference job.

Gap 2

The period between the last time in high school full-time and the beginning of the job held in the week preceding the interview (or the beginning of the week preceding the interview for those who did not have a job last week).

Graduates

Respondents who have graduated from high school (Grades 11, 12 or 13 depending on the province) are those who had completed the requirements for a high school graduation certificate, diploma or its equivalent. Respondents to the survey may have considered themselves to be a graduate if they received a certificate after completing Grade 10. It should be noted that these data were self-reported and not verified.

High School/Secondary School

The educational structure varies across the provinces. In general, at the secondary (high school) level, there is usually a choice of at least two programs: academic or vocational. Some secondary schools may specialize in mainly vocational training (technical and commercial) but most secondary schools offer both academic courses (preparatory to university) and vocational courses, which prepare students either for an occupation or for further postsecondary non-university education.

Household

Refers to a person or group of persons (usually this does not include foreign residents), who occupy the same dwelling and do not have a usual place of residence elsewhere in Canada. It may consist of a family group (census family) with or without other non-family persons, of two or more families sharing a dwelling, of a group of unrelated persons, or of one person living alone.

Career or Job Related Education or Training

Refers to any education or training activities taken for the development or upgrading of skills to be used in a present or future career/employment position.

Junior High/Middle School

Usually refers to instruction at the grade 7, 8 or 9 level and in some cases grade 10.

Province	Junior High Grades
Newfoundland, Prince Edward Island, Nova Scotia, New Brunswick, Saskatchewan, Alberta, Northwest Territories	7,8,9
Quebec (English schools)	Sec. I, II
Quebec (French schools), Manitoba	not applicable
Ontario	7, 8, 9, 10
British Columbia, Yukon	8, 9, 10

Labour Force Status

Respondents (excluding 1991 continuers in the SLFS) were classified as employed, unemployed, or not in the labour force. Employed respondents are those who have a job and are not on layoff or awaiting recall. Those not working and looking for work, on temporary layoff or awaiting recall, awaiting the start of a job were unemployed. Respondents not in the labour force were not working nor looking for work. Full-time students looking for full-time work were also classified as not in the labour force.

Main Job Last Week

Is the job or business that respondents spent the most hours at last week.

National Occupational Classification Codes (NOC)

The 1992 National Occupational Classification Coding system is designed to provide a systematic classification structure to identify and categorize the entire range of occupational activity in Canada. Information on occupations gathered in the 1995 Follow-up Survey is coded using the NOC and information on the last job held during the reference week in the 1991 School Leavers Survey has been re-coded using the NOC.

Paid 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 where unpaid family work is defined as unpaid work which contributed directly to the operation of a farm, business, or professional practice owned or operated by a related member of the household. Unpaid family work does not include persons who stay at home to manage the household.

Postsecondary programs

1) University programs leading to bachelor's, master's or doctoral degrees, or specialized certificates or diplomas.
2) Programs offered at CAATs, CEGEPs, community colleges, technical schools, hospital schools of nursing and similar institutions. Included are programs of one year duration or longer which normally require secondary school completion or its equivalent for admission.

Pre-vocational Programs

Pre-vocational programs provide students with the prerequisites needed to enter a trade/vocational or postsecondary program. The programs are aimed at improving a candidate's knowledge in the basic subjects of Mathematics, English or French, and General Science. Completion of these courses is not necessarily equivalent to high school graduation.

Private Elementary/Secondary Schools

These schools, whether church-affiliated or non-sectarian, are operated and administered by private individuals or groups.

Private Employment Agency

Privately-owned businesses that supply labour on demand to other businesses, generally on a short-term or temporary basis.

Private Training Institution

Privately-owned schools licensed by the province which are profit-oriented and are engaged in providing professional

and vocational training.

Province of Study

Refers to the province in which the respondent was last in high school, junior high or elementary school.

Reference Job

A reference job is a job at which a respondent worked for the same **employer** 20 hours or more per week, for at least 6 months in a row. This includes self-employment.

Response Rate

This number evaluates a data collection effort. It is the number of units responding to the survey, divided by the total number of units in the sample. The denominator includes all units in the sample population who were selected but did not respond for whatever reason.

Returners

Respondents who had ever dropped out of school and had returned to school after dropping out. Included are non-completers who had dropped out of school and returned again before finally dropping out, high school graduates who had ever dropped out and continuers who had ever dropped out.

Sample Population

A selection (usually through random selection) of units from an entire population. The randomly selected sample completes the survey questionnaire. Inferences are made from the sample population to the target population. It is important, therefore, that the sample population be representative of the target population.

School Leavers

Respondents who have not graduated from high school and are not attending elementary or secondary school.

Sector: General, Academic, Long Vocational, Short Vocational (Quebec)

Long Vocational:

After reaching age 16 and upon getting a secondary diploma a student may enter a one or two year vocational program. Students completing Secondary IV credits in his/her mother tongue, second language, mathematics and moral or religious education may also enter. The program leads to a “Diplôme d’études professionnelle” (DEP), diploma of vocational education. After obtaining a DEP, a student may follow a program of more specialized vocational training leading to “Attestation de specialisation professionnelle” (ASP). Before 1987-88, long vocational programs were given in Secondary IV and V.

Short Vocational:

A student may enter high school vocational after getting Secondary III credits in his/her mother tongue, a second language and mathematics. This training leads to a “Certificat d’études professionnelle” (CEP), certificate of vocational education after 2 years. Before 1987-88 the short vocational sector referred to vocational training received in Secondary III and IV.

Social Assistance, Welfare

Money, vouchers and services given by provincial or municipal authorities to those who are very badly off. A needs test is required to qualify for social assistance. Includes aid given to persons with mental or physical limitations.

Standard Industrial Classification Codes (SIC)

The 1980 Standard Industrial Classification system is designed to provide a systematic classification structure to identify and categorize the entire range of industries in Canada. The SIC is used to classify industry information in the 1995 Follow-up Survey.

Standard Occupational Classification Codes (SOC)

The 1980 Standard Occupational Classification system is designed to provide a systematic classification structure to identify and categorize the entire range of occupational activity in Canada. The SOC is used to classify job information in the 1991 School Leavers Survey.

Permanent/Temporary Job

A permanent job is one for which there was no indication that the job would end at some definite point in time eg. hired permanently with no specified term. A temporary job is one for which there was a definite indication that the job will terminate at some specified point in time, eg. hired for a six month term.

Trade/vocational Training or Education

Activities and programs that provide the skills needed to function in a particular vocation. These programs emphasize manipulative skills and well-defined or well-established procedures, rather than the application of ideas and principles.

Transitional School

A school that provides training to allow students to function effectively in the labour market. This training is usually given to older students or to students who have had previous difficulties in a regular school.

University

Independent institution granting degrees in at least arts and sciences.

Vocational or Trade School

Technical and trades training varies between and within provinces. It is offered in both public and private institutions such as community colleges, institutes of technology, trade schools and business colleges. It may also take place on the job, in apprenticeship programs or in industry training programs.

Work Experience Program

Educational programs are those which combine high school study with on-the-job training.

5. Survey Methodology

5.1 Target Population and Sampling Frame

The sampling frame for the survey was files from the federal Family Allowance (FA) program, which was considered to be the most complete listing of young people in Canada available at the time of the SLS. (It should be noted that the Family Allowance program was discontinued in 1993.) The FA files also provided indicators such as "child has left home", "child is working", etc., which were used to create a payment status variable to identify potential leavers, helping to ensure an adequate number of leavers for analysis. This payment status variable indentified respondents as "potential" or "non-potential" leavers. The "potential" leavers and "non-potential" leavers were sampled at different rates with the rate in general higher for "potential" leavers. Several years of FA files were linked by an identification number that does not change as long as the individual remains within the same province. Because the same identification number is not used when an individual moves to a different province, matching between provincial files also occurred. Matching of records was also required if the support of the individual was transferred to the other parent or to an agency. After the required linking, the sampling frame consisted of five years of Family Allowance files, from 1986 to 1990 (Social Survey Methods Division).

The SLS target population consisted of young people who were 18 to 20 years old as of April 1st, 1991, and residing in one of the 10 provinces in Canada. Excluded were residents of the Yukon and the Northwest Territories, as well as young people whose last location of residence was listed as an agency (e.g., orphanage, prison, transition home, shelter, etc.).

The age range of 18 to 20 years old inclusive was chosen to allow both evaluation of the leaver rates and comparison of the three groups of school attendees, namely leavers, graduates and continuers. The national and provincial leaver rates would be evaluated using the sample of 20-year-olds, as youth of this age were more likely to be either graduates or leavers, rather than the temporary state of continuers.

The timing of the SLFS was consistent with the belief that four years after the initial survey would be an appropriate time to re-contact these individuals, since by 1995 they would be 22, 23 or 24 years old and likely would have had one or more jobs. In addition, most continuers in 1991 would be either graduates or leavers, allowing a more detailed labour market analysis. In contrast, most SLS respondents had limited labour market experience and many were excluded from the labour market analysis because they were still attending high school.

5.2 Sample Design

5.2.1 Stratification

The SLS sample was selected from the frame described above using a stratified sampling design, with strata defined by age (on April 1, 1991), FA province of residence and payment status, a variable generated from information codes found in the Family Allowance file to aid in the selection of individuals who have dropped out of secondary school. Each age and province of residence cell was subdivided using payment status to create a further breakdown to potential and non-potential leavers. Population counts by stratum are given in Table 1.

Table 1: Population Counts by Stratum

Province	Age 18	Age 18	Age 19	Age 19	Age 20	Age 20	Total
	Potential	Non-Potential	Potential	Non-Potential	Potential	Non-Potential	
Newfoundland	111	11,424	172	11,910	181	11,703	35,501
P.E.I.	25	2,003	32	2,195	54	2,122	6,431
Nova Scotia	132	13,576	163	14,167	175	14,357	42,570
New Brunswick	185	11,998	270	12,261	302	12,281	37,297
Quebec	557	86,514	1,096	88,579	1,046	91,928	269,720
Ontario	1,618	135,231	1,922	137,478	2,006	142,403	420,658
Manitoba	144	16,174	200	16,413	215	16,939	50,085
Saskatchewan	277	15,283	329	15,392	356	15,563	47,200
Alberta	360	34,801	499	35,567	644	36,919	108,790
British Columbia	443	40,349	460	41,159	559	44,056	127,026
Total	3,852	367,353	5,143	375,121	5,538	388,271	1,145,278

5.2.2 Sample Size and Selection

The SLS sample consisted of 18,000 individuals from across Canada, excluding Yukon and Northwest Territories. The sample was selected to guarantee national and provincial leaver rates for 20-year-olds within a coefficient of variation of 16.5%, and to have continuers, leavers and graduates, each considered separately, possess some characteristic from the questionnaire which could be estimated within a coefficient of variation of 16.5%. A simple random sample was taken from each stratum except when it was required to take all in a stratum. Sample counts by stratum for the SLS are given in Table 2.

Table 2: SLS Sample Counts by Stratum

Province	Age 18 Potential	Age 18 Non-Potential	Age 19 Potential	Age 19 Non-Potential	Age 20 Potential	Age 20 Non-Potential	Total
Newfoundland	111	209	172	173	181	469	1,315
P.E.I.	25	105	32	108	54	566	890
Nova Scotia	132	228	163	217	175	475	1,390
New Brunswick	185	165	270	105	302	368	1,395
Quebec	557	353	575	415	330	370	2,600
Ontario	865	480	742	523	330	370	3,310
Manitoba	144	241	200	205	215	435	1,440
Saskatchewan	277	163	253	177	356	344	1,570
Alberta	360	290	380	270	330	370	2,000
British Columbia	443	247	411	289	330	370	2,090
Total	3,099	2,481	3,198	2,482	2,603	4,137	18,000

All SLS respondents were in-sample for the Follow-up survey, with the exception of 18 individuals who participated in a SLFS pre-test and 11 who indicated in 1991 that they did not want to participate in further surveys. Sample counts for the SLFS are given in Table 3. The counts are given by Province of Interview (in the 1991 survey) and Year of Birth as provided by the respondent in the SLS. These values may differ from the Family Allowance last province of residence and year of birth.

Table 3: SLFS Sample Counts

Province of Interview	Year of Birth					Total
	1969	1970	1971	1972	1973	
Newfoundland		256	280	167	39	742
P.E.I.		259	147	84	15	505
Nova Scotia		258	260	182	57	757
New Brunswick		296	251	172	39	758
Quebec	2	291	488	459	100	1,340
Ontario	1	323	626	686	154	1,790
Manitoba		260	244	208	42	754
Saskatchewan		232	189	199	41	661
Alberta	2	276	355	352	71	1,056
British Columbia		274	352	358	73	1,057
Yukon Territory		2	2			4
N.W. Territories		5	1			6
Other				1		1
Total	5	2,732	3,195	2,868	631	9,431

6. Data Collection

6.1 Sampling Frame

The sampling frame for the survey was the Family Allowance files. These files were decided on as the sampling frame because potential leavers could be pre-identified to help ensure an adequate number of leavers for analysis. These files were also considered to be the most complete listing of young people in Canada available at the time of the SLS. In order to create this frame records from year to year were linked by an identification number that does not change as long as the individual remains within the same province. Because the same identification number is not used when an individual moves to a different province matching between provincial files also occurred. Matching of records was also required if the support of the individual was transferred to the other parent or to an agency. After the required linking, the sample frame consisted of five years of Family Allowance Files, 1986 to 1990.

6.2 Target Population

After a test conducted in 1990, it was decided that the ages of 18, 19 and 20 would be best suited to answer the required objectives of the SLS. Four years after the initial survey was deemed an appropriate time to re-contact these individuals, since they would be 22, 23 and 24 years old and likely would have had one or more jobs. In addition, most continuers in 1991 would be either graduates or leavers, allowing a more detailed labour market analysis. In contrast, most SLS respondents had limited labour market experience and many were excluded from the labour market analysis because they were still attending high school.

The Follow-up survey was conducted between September and December of 1995, allowing information to be collected on school and work activity (including summer employment) simultaneously. Also in the fall it was less difficult to identify future plans of youth, such as going back to school (continuers), graduated and working or continuing on in post-secondary.

6.2.1 Data Collection Method - SLS

The SLS was conducted by telephone using a Computer Assisted Telephone Interviewing (CATI) system. Due to the length and complexity of the questionnaire, mail-out surveys were not conducted. Problems of low response rates and difficulty in self-enumeration that would be associated with a mail-out survey also led to the decision to use a CATI system. Personal interviews were deemed too expensive. Also, once the respondent was contacted using the phone, which was believed to be the most efficient tracing method, it seemed most economical to complete the questionnaire over the telephone.

For the SLS, telephone billing files were matched to the addresses on the sampling frame to provide the interviewers with several numbers from which they could attempt to trace the individual. If this method proved unsuccessful, interviewers tried other techniques such as phoning neighbours and contacting directory assistance. For 34 percent of the sample there was no success in finding either the individual or the parent/guardian, while for 6 percent for the sample the parent/guardian was traced but the individual was not traced. Of the 10,792 youths that were traced, 9,460 were completed responses.

6.2.2 Data Collection - SLFS

In the SLS, respondents were asked to provide an address and telephone number where they could be reached if Statistics Canada wanted to contact them for a Follow-up. They were also asked to provide the name, address and

telephone number of a friend, relative or neighbour who could be contacted if the individual moves. If the respondent could not be located using the information from the 1991 SLS, the individual was placed in the "tracing" module, which linked the 1991 name and address information to 1995 telephone files.

If this failed to produce a correct phone number, interviewers were asked to use conventional tracing methods such as directory assistance. Completed responses were obtained for 6,284 of the 9,460 individuals that were traced.

6.3 Tracing and Response Rates for the School Leavers and Follow-Up Surveys

Interviewers are instructed to make all reasonable attempts to obtain interviews with respondents. For individuals who at first refuse to participate in the SLS, interviewers stress the importance of the survey and the respondent's cooperation. For cases in which the timing of the interviewer's call is inconvenient, an appointment is arranged to call back at a more convenient time. For cases in which there is no one home, numerous call backs are made. Under no circumstances are sampled youths replaced by other youths for reasons of non-response.

The contact point available on the frame (Family Allowance file) was the recipient's address, the recipient being one of the parents or guardians of the respondent to be contacted. The following table shows the tracing and response rates at the Canada level:

TRACING AND RESPONSE RATES - SLS

	Non-response	Out of Scope	Responses	TOTAL
Untraced Recipient Untraced Respondent	6,065 (34%)	-	-	6,065 (34%)
Traced Recipient Untraced Respondent	1,134 (6%)	19 (0%)	-	1,153 (6%)
Traced Recipient Traced Respondent	1,151 (6%)	171 (1%)	9,460 (53%)	10,782 (60%)
TOTAL	8,350 (46%)	190 (1%)	9,460 (53%)	18,000 (100%)

The next stage of data collection was to administer the SLS questionnaire to the youth. As you can observe in the table above, the response rate once the youth is traced is high: of the 10,782 youths traced, 9,460 youths were complete respondents and in-scope to the survey (88%).

An analogous table for the 1995 SLFS is given below. Unlike the SLS, in the SLFS there is no distinction between "Recipient" and "Respondent", although a distinction is made between cases where a phone number was found but it was not possible to verify that the phone number belonged to the respondent (the middle line of the table), and cases where the respondent was contacted but was unable to or chose not to complete the survey (Traced respondent, Non-response).

TRACING AND RESPONSE RATES - SLFS

	Non-response	Deceased	Responses	TOTAL
Untraced Respondent (no phone # found)	1,354 (14.3%)	-	-	1,354 (14.3%)
Phone # found but identity not verified	873 (9.2%)	-	-	873 (9.2%)
Traced Respondent (identity verified)	924 (9.8%)	25 (0.3%)	6,284 (66.4%)	7,233 (76.5%)
TOTAL	3,151 (33.3%)	25 (0.3%)	6,284 (66.4%)	9,460 (100%)

Of the 7,233 SLS respondents who were successfully traced for the SLFS, complete responses (including agreement to data sharing) were obtained for 6,284 (86.9%).

6.4 Implementation by Wave

The SLS was administered in a series of five waves to ensure that all respondents had an equal chance of being contacted, with each wave to be in the CATI system for the same length of time. The cases that had not been contacted after each of the first four waves then became the fifth wave and attempts to contact were made again. No similar implementation by wave was done for the SLFS.

6.5 Corrections for Potential Bias

The only adjustment performed for the SLS was to post-stratify by sex, although other potential sources of bias such as time zone differences, implementation by wave and number of calls made to the respondent were considered and investigated.

During the administration of the SLFS several steps were taken to correct for potential bias. In order to ensure an adequate number of leavers for analysis, provincial target response rates for leavers were set and emphasis was directed where appropriate to meet these targets. Provincial response rates were monitored daily.

The weighting procedures described below contain features which will correct for potential bias, and in addition "Not stated" categories were created for some items as a means of addressing item non-response. It would take months to calculate any potential bias and would most likely not have any major impacts on the results of the SLFS. A special project to investigate this may be initiated if time and cost allow.

7. Data Processing

In addition to the publications that have been produced from the "School Leavers Survey", the major output is a "clean" microdata file. This section presents a brief summary of the processing steps involved in producing this file.

7.1 Data Capture

By using a CATI system to collect the information in both the SLS and SLFS, a computer file was generated as the survey progressed. This file was saved at the end of each day and then appended to a completed document file. Many of the logical edits that would normally be performed on the document after collection were programmed into the collection instrument, thereby saving time and editing problems.

7.2 Editing and Imputation

1991 SLS

The first type of errors treated were errors in questionnaire flow, where questions which did not apply to the respondent 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. This only occurred when the interviewer had to "back up" during the interview to change a response which then put the respondent on a different flow through the questionnaire. In most cases the CATI system ensured that the correct sequence of questions were answered.

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

1995 SLFS

Survey processing for the SLFS involved editing all remaining survey records according to pre-specified edit rules to check for errors, gaps and inconsistencies in the survey data. Checks were made to ensure that numerical answers to certain questions fell within acceptable logical ranges. Checks were also made to ensure that portions of the questionnaire that were skipped in the interview because of a previous answer were, in fact skipped. Where errors or inconsistencies were found, the erroneous information was either blanked out or replaced by a "not stated" or "don't know" response. When appropriate, imputation stategies were used to assign a value for missing data.

7.3 Coding of Open-ended Questions

A few data items on the questionnaire were recorded by interviewers in an open-ended format. Following the data collection process, a team of specialized coders was given the task of looking at the written responses and coding them according to existing standard education coding lists of courses, standard occupational classification codes, national occupational classification codes and/or standard industrial classification codes.

For those questions where there were no existing lists, responses to the open-ended questions, such as "other - please specify", were coded by a subject matter analyst.

7.4 Creation of Derived Variables

Certain items on the microdata files were combined to form derived variables. Among them are the type of student, labour force status, age of event, highest grade completed in primary/secondary school, ever left school indicator and so on.

8. Weighting

The principle behind estimation in a probability sample such as the SLS and the SLFS 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, the number of individuals in the population represented by it. This number is known as the weight, and the SLFS weight appears on the SLFS microdata file as the variable SLFWGT. This weight must be used to derive estimates from the SLFS microdata file. For example, if the number of individuals who ever left school is to be estimated, it is done by selecting the records of individuals in the sample who left school and summing the weights assigned to those records.

Because the SLFS is a Follow-up to the SLS, the SLFS weight is a function of the SLS weight, and consequently a summary of the SLS weighting will be given before describing the SLFS weighting. It should be noted that the SLS weight does not appear on the SLFS microdata file.

As the weight is a function of both the population size and number of respondents, the SLS weight and SLFS weight assigned to a respondent may be different, with the SLFS weight in general larger. This follows from the fact that the SLS and SLFS samples both represent the same target population, but since the SLFS sample is smaller, each record is required to represent more individuals and therefore carries a larger weight.

8.1 SLS Weighting

SLS weighting classes were defined by four variables, namely the three stratification variables (province, age and potential dropout indicator) and sex. All respondents in the same weighting class receive the same weight.

The final SLS weight attached to each record was the product of the following factors: a basic sampling weight (i.e. the inverse of the probability of selection), an adjustment to account for non-response to the SLS, and an adjustment for sex. Each is described below:

Basic Weight

The basic weight of a record is merely 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 basic weight for each record would be 1/0.02=50.

As mentioned earlier, the SLS sample was selected using a stratified design, with strata defined by province, age and potential dropout indicator. Although the first name of the youth appeared on the Family Allowance files, the gender of the youth did not, and consequently sex could not be used as a stratification variable. Consequently the basic weight is a function of only the three stratification variables and is not a function of sex.

Adjustment for Non-response

Notwithstanding controls, some non-response is inevitable, despite all attempts made by interviewers. Non-response is compensated for by proportionally increasing the weights of responding youths. The weight of each responding record is increased by the ratio of the number of youths that should have been interviewed to the number that were actually interviewed. This adjustment was done separately for each stratum, and is based on the assumption that the youths that have been interviewed represent the characteristics of those who should have been interviewed. To the extent that this assumption is not true, estimates will be somewhat biased.

Adjustment for Sex

The gender of respondents was collected during the interview, and for each first name the male/female breakdown obtained in the survey was used to give an imputation strategy for sex. For instance, if there were, say, 100 SLS respondents named Pat, of whom 60 were female and 40 were male, then in the imputation strategy the gender variable of an individual named Pat would be assigned “male” with probability 0.6, and “female” with probability 0.4. Application of this imputation strategy to non-respondents and non-sampled individuals on the frame, along with the gender of respondents collected in the survey, provided weighting-class population totals, and for each weighting class an adjustment to account for differences between males and females was formed as the ratio of this population total to the sum of the sex-adjusted weights of all respondents in the weighting class.

The final SLS weight assigned to each respondent was the product of the three factors above, and for each weighting class this product simplifies to the ratio of the population total to the number of SLS respondents. It should be noted that the SLS weight does not appear on the SLFS microdata file.

8.2 SLFS Weighting

SLFS weighting classes were defined by the four variables which defined SLS weighting classes (namely province, age, potential dropout indicator and sex) as well as SLS type (continuer, leaver, graduate). Certain groups of weighting classes were collapsed (by potential dropout indicator or by potential dropout indicator and sex) in order to allow no weighting class with less than two respondents.

Statistics Canada's Generalized Estimation System (GES) was used to produce SLFS weights. An estimate of the weighting class population total was calculated as the sum of the SLS weights of all SLS respondents in the weighting class. For each weighting class, the SLFS weight was calculated as the ratio of this population total estimate to the number of SLFS respondents. Maximum, minimum and mean SLFS weights by province are presented in Table 4 below.

Table 4: SLFS Weights by Province

Province	Minimum Weight	Mean Weight	Maximum Weight
Newfoundland	1.68	62.65	225.60
Prince Edward Island	2.50	16.51	52.51
Nova Scotia	3.32	79.61	165.93
New Brunswick	2.29	68.40	334.54
Quebec	3.11	290.10	943.34
Ontario	6.28	386.44	1550.25
Manitoba	2.29	98.30	327.70
Saskatchewan	3.98	98.05	291.08
Alberta	2.93	167.54	580.20
British Columbia	3.15	198.56	944.28

The SLFS weighting strategy described above produces SLFS weights which are "benchmarked" to SLS population estimates, leading to consistent estimates using the 1991 and 1995 survey data for each of 1991 graduates, continuers and leavers. Minor discrepancies between SLS and SLFS estimates by sex may be introduced by the collapsing of sex in forming some SLFS weighting classes.

9. Data Quality

9.1 Sampling Error

The SLFS produces estimates based on information collected from and about a sample of individuals. 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 estimates.

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 such as this, 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 may be expressed as a percentage.

For example, suppose that based upon the SLFS survey results, one estimates the proportion of the survey population which does not plan to take some further training or education in the next 5 years to be 0.133, and that this estimate is found to have standard error of 0.00681. Then the coefficient of variation of the estimate is calculated as:

$$\frac{0.00681}{0.133} * 100\% = 5.1\%$$

9.2 Non-sampling Error

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 error.

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. Quality assurance measures were implemented at each step of the data collection and processing cycle to monitor the quality of the data. These measures included the use of highly-skilled interviewers, training of interviewers with respect to the SLFS 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.

9.2.1 Total Non-response

Total non-response is a major source of non-sampling error if non-respondents cannot be represented by respondents. In the SLFS total non-response occurred because the interviewer was either unable to contact the respondent or the respondent refused to or was unable to participate in the survey. Total non-response was addressed by adjusting the

weight of those who responded to compensate for those who did not respond.

9.2.2 Partial Non-response

Partial non-response in the SLFS may have occurred if the respondent either refused to answer a question, did not understand a question, or could not recall the requested information. Overall, it is unlikely that partial non-response contributed significantly to non-sampling error in the SLFS. One measure taken to address partial non-response was to create "Don't know" categories for many items.

It should be noted that in the course of the SLFS interview the respondent was asked to recall a number of dates, often of events which occurred many years in the past. Although these dates were subjected to a series of edits to help ensure consistency, the dates appearing on the file are still greatly influenced by the ability of the respondent to recall, and are subject to the limitations of this capacity.

9.2.3 Variable Notes

Flag1 (reference job updates) -

For a small number of cases, the dates of the first reference job and the most recent reference job appear to be inconsistent. These respondents used the end dates of these jobs to assess the first and most recent. That is, the most recent reference job was the one that ended most recently.Cases where this situation applies are indicated with a flag on the data file. This flag is found in field 404.

Volunteer Work (Q217) -

The percentage of respondents doing any volunteer work in the last twelve months preceding the SLFS appears high compared to previous data. However comparative data on volunteer activities for a similar age cohort is only available for 1987. Caution should be exercised with this variable as it appears that respondents may have used a broad definition of volunteer activity.

9.2.4 Corrections for Potential Bias

During the administration of the SLFS several steps were taken to correct for potential bias. In order to ensure an adequate number of leavers for analysis, provincial target response rates for leavers were set and emphasis was directed where appropriate to meet these targets. Response rates were monitored daily.

It would take months to calculate any potential bias and would most likely not have any major impact on the results of the SLFS. A special project to investigate this may be initiated if time and cost allow.

10. Guidelines For Tabulation, Analysis, Publication And Release

It is important for users to become familiar with the contents of this section before publishing or otherwise releasing any estimates derived from the microdata file of the School Leavers Follow-up Survey.

This section of the documentation outlines the guidelines to be adhered to by users publishing or otherwise releasing any data derived from the survey microdata file. With the aid of these guidelines, users of the microdata file 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.

This section consists of four sub-sections - rounding guidelines, sample weighting guidelines, sampling variability guidelines (i.e. CV release guidelines) and guidelines for statistical analysis.

10.1 Rounding Guidelines

In order to ensure that estimates for publication or other release derived from the SLFS microdata file will 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.

10.2 Sample Weighting Guidelines for Tabulation

The sample design used for the School Leavers Follow-up Survey was not self-weighting. When producing simple estimates, including the production of ordinary statistical tables, users must apply the sampling weights placed on the individual records on the microdata file. Otherwise, the estimates derived from the microdata file 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, because of their treatment of the weight field, may not allow the generation of estimates that match exactly those available from Statistics Canada.

Before discussing how these measures can be obtained 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 School Leavers Follow-up Survey.

10.2.1 Definition of Categorical and Quantitative Estimates

Categorical Estimates

Categorical estimates are estimates of the number, proportion or percentage of the surveyed population possessing certain characteristics or falling into some defined category. The number of individuals who left school because they were bored or the proportion of individuals who left school because they were bored are examples of such estimates. An estimate of the number of persons possessing a certain characteristic or falling into some defined category is referred to as an estimate of an **aggregate**.

Quantitative Estimates

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

An example of a quantitative estimate is the average number of months between the first and last time a non-completer dropped out. The numerator is an estimate of the total number of months reported and the denominator is an estimate of the number of individuals who didn't complete primary/secondary school and dropped out more than once.

10.2.2 Tabulation of Categorical Estimates

Estimates of the number of individuals with a certain characteristic can be obtained from the microdata file by summing the weights of all records possessing the characteristic of interest. Proportions and ratios of the form $\hat{X}/\hat{Y}$ are obtained by:

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

10.2.3 Tabulation of Quantitative Estimates

Estimates of quantities can be obtained from the microdata file by multiplying, for each record, the value of the variable of interest by the weight, then summing this quantity over all records of interest.

To obtain a weighted average of the form $\frac{\sum x_i w_i}{\sum w_i}$, the numerator ($\sum x_i w_i$) is calculated as for a quantitative estimate and the denominator ($\sum w_i$) is calculated as for a categorical estimate.

10.3 Sampling Variability Guidelines

Before releasing and/or publishing any estimate from the SLFS microdata file, users should first determine the number of respondents who contribute to the calculation of the estimate. If this number is less than 30, the weighted estimate should not be released regardless of the value of the coefficient of variation for this estimate. For weighted estimates based on sample sizes of 30 or more, users should first determine the coefficient of variation of the rounded estimate (see Section 11) and follow the guidelines below.

Sampling Variability Guidelines for Estimates Produced from
The School Leavers Follow-up Survey Microdata File

Type of estimate	CV (in %)	Guidelines
1. Unqualified	0.0 -16.5	Estimates can be considered for general unrestricted release. Requires no special notation.
2. Qualified	16.6 - 25.0	Estimates can be considered for general unrestricted release but should be accompanied by a warning cautioning subsequent users of the high sampling variability associated with the estimates. Such estimates should be identified by the letter Q (or in some other similar fashion).
3. Restricted	25.1 - 33.3	Estimates can be considered for general unrestricted release only when sampling variabilities are obtained using an exact variance calculation procedure. Unless such variances are obtained, such estimates should be deleted and replaced by dashes (---) in statistical tables.
4. Not for release	33.4 or greater	Estimates cannot be released in any form under any release OR circumstances. In statistical tables such estimates should be deleted and replaced by dashes (---).

10.3.1 Release Cut-offs for the SLFS

As an aid in the application of the CV release guidelines presented above, Table 5 below gives minimum estimate sizes (i.e. minimum estimated number of individuals possessing the characteristic) to attain certain coefficients of variation, for domains defined by Province of study in 1995, by Gender and by Type of person in 1995. The minimum estimate size is called the "cut-off" number, and cut-offs to attain a coefficient of variation of 16.5%, 25% and 33.3% are given in Table 5. From the CV release guidelines, these three coefficients of variation correspond to the minimum standard for "Unqualified", "Qualified" and "Restricted" estimates, respectively. A user can determine in which range the CV will fall using the cut-off numbers, although either an exact variance or an approximate CV (using the Approximate Sampling Variability Tables as described in Section 11) still remains to be calculated.

Table 5: Release Cut-offs for Totals (i.e. Minimum Estimate Sizes)

Domain	Minimum Estimate Size		
	CV=16.5%	CV=25%	CV=33.3%
Canada	17,000	7,500	4,000
Newfoundland	3,500	1,500	1,000
Prince Edward Island	1,000	500	500
Nova Scotia	3,500	1,500	1,000
New Brunswick	5,000	2,500	1,500
Quebec	19,000	8,500	5,000
Ontario	24,000	11,000	6,000
Manitoba	5,000	2,500	1,500
Saskatchewan	5,000	2,500	1,500
Alberta	9,000	4,000	2,500
British Columbia	11,500	5,000	3,000
Males	17,500	7,500	4,500
Females	16,500	7,500	4,000
Graduates	17,000	7,500	4,000
Continuers	5,500	4,000	3,500
Leavers	20,500	9,500	5,500

For example, suppose a user wishes to estimate the number of individuals in Ontario who took various types of training, using Q35A and Q36 of the SLFS microdata file. Suppose the SLFS microdata file gives that 2,000 individuals completed training towards a "first professional" degree, 9,500 individuals towards a private business or commercial school diploma or certificate, 22,500 individuals towards a trade/vocational or registered apprenticeship diploma or certificate, and 135,000 individuals towards a college or CEGEP diploma or certificate. Using Table 5, the CV of the first estimate exceeds 33.3%, the CV of the second estimate lies between 25% and 33.3%, the CV of the third estimate lies between 16.5% and 25%, and the CV of the fourth estimate is less than 16.5%. (In Example 2 in Section 11.1.1, the approximate CV of the fourth estimate from the ASVT is found to be 6.2%.)

10.4 Guidelines for Statistical Analysis

The SLFS sample design incorporates stratification and unequal probabilities of selection of respondents. Using data from such a complex survey 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 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 almost meaningless.

For many analysis techniques (for example, linear regression, logistic regression, analysis of variance), a method exists which can make the application of standard packages more meaningful. If the weights on the records are rescaled so that the average weight is one (1), then the results produced by the standard packages will be more reasonable. They still will not take into account the stratification of the sample's design, but they will take into account the unequal probabilities of selection. The rescaling can be accomplished by dividing each weight by the overall average weight before the analysis is conducted.

Approximate variances for simple estimates such as totals, proportions and ratios (i.e. for qualitative variables) are described in Section 11 and are calculated using the Approximate Sampling Variability Tables provided in Appendix D.

11. Approximate Sampling Variability Tables

In order to supply coefficients of variation which would be applicable to a wide variety of categorical estimates produced from the SLFS microdata file and which could be readily accessed by the user, a set of Approximate Sampling Variability Tables (ASVT) has been produced. These "look-up" tables, which can be found in Appendix D, allow the user to obtain an approximate coefficient of variation based on the size of the estimate calculated from the survey data.

These coefficients of variation (CV) are derived using the variance formula for simple random sampling and incorporating a factor which reflects the stratified nature of the SLS sample design. This factor, known as the design effect, has been determined by first calculating design effects for a wide range of characteristics and then choosing from among these a conservative value to be used in the look-up tables and which would then apply to the entire set of characteristics. It is to be noted that all coefficients of variation in these tables are approximate and therefore unofficial, and the corresponding exact variance must be calculated in order to make definite conclusions regarding the CV.

Design effects were calculated and an ASVT produced for each of the following 42 levels of estimation:

- Canada (1)
- Province of study in 1995 (10)
- Gender (2)
- Type of person in 1995 (3)
- Gender by Type of person in 1995 (6)
- Province of study in 1995 by Type of person in 1995[2] (20)

Table 6 gives the design effect and respondent and population counts for the levels defined above.

Table 6: Design Effects and Respondent and Population Counts

Level of Estimation	Design Effect	Respondent Count	Population Count
Canada	2.57	6,284	1,136,237
Newfoundland--Graduates	1.42	388	28,031
Newfoundland--Leavers	3.69	171	6,958
Newfoundland	1.62	565	35,338
Prince Edward Island--Graduates	1.86	295	5,127
Prince Edward Island--Leavers	2.59	90	1,226
Prince Edward Island	1.79	386	6,396
Nova Scotia--Graduates	1.25	417	35,721
Nova Scotia--Leavers	2.52	112	6,520
Nova Scotia	1.32	530	42,246
New Brunswick--Graduates	1.79	348	31,624

[2]The small number of 1995 continuers prohibited production of provincial ASVT for 1995 continuers.

Table 6: Design Effects and Respondent and Population Counts

Level of Estimation	Design Effect	Respondent Count	Population Count
New Brunswick--Leavers	8.36	185	4,579
New Brunswick	2.23	540	36,940
Quebec--Graduates	1.38	567	216,887
Quebec--Leavers	5.51	336	47,029
Quebec	1.93	920	267,531
Ontario--Graduates	1.39	741	367,720
Ontario--Leavers	6.45	319	49,024
Ontario	1.79	1,083	418,085
Manitoba--Graduates	1.43	393	41,894
Manitoba--Leavers	3.22	114	7,941
Manitoba	1.51	509	49,949
Saskatchewan--Graduates	1.52	371	40,698
Saskatchewan--Leavers	4.01	98	5,740
Saskatchewan	1.63	473	46,707
Alberta--Graduates	1.32	449	92,572
Alberta--Leavers	4.34	185	14,887
Alberta	1.56	640	107,752
British Columbia--Graduates	1.35	460	108,284
British Columbia--Leavers	5.66	175	16,986
British Columbia	1.74	638	125,292
Males--Graduates	2.01	2,017	470,874
Males--Continuers	4.27	30	5,234
Males--Leavers	6.63	975	103,577
Males	2.53	3,022	579,685
Females--Graduates	2.24	2,412	497,685
Females--Continuers	7.25	40	1,554
Females--Leavers	9.30	810	57,313
Females	2.72	3,262	556,552
Graduates	2.14	4,429	968,559
Continuers	7.19	70	6,788
Leavers	7.11	1,785	160,890

It should be noted that the design effect for levels involving 1995 leavers and continuers are higher than for the corresponding level involving graduates.

Estimates of the actual variance for specific variables may be obtained from Statistics Canada on a cost recovery basis. The use of actual variance estimates would likely result in estimates with lower variances; for example, estimates in the “Not for release” category according to the ASVT could move up to the “Restricted” category. See Section 10.3 for more information on CV release guidelines.

11.1 How to use the ASVT for Categorical Estimates

The following rules should enable the user to determine the approximate CV from the ASVT for estimates of the number, proportion or percentage possessing a certain characteristic and for ratios and differences of such estimates, for both the entire surveyed population and for subsets (domains) of the surveyed population.

When estimating for a certain domain, it is necessary to refer to the appropriate ASVT. For domains which do not coincide with one of the 42 levels of estimation described above, the appropriate ASVT is that of a level of estimation which includes the domain. For example, for a domain formed of female 1995 graduates, refer to the ASVT for "Females - Graduates", whereas for a domain formed of females in New Brunswick, refer to either the "Females" or "New Brunswick" tables.

The actual tables are found in Appendix D.

Rule 1: **Estimate of number possessing a characteristic (aggregates)**

On the appropriate ASVT, locate the number in the left-most column of the table (headed "Numerator of Percentage") which is closest to the estimate of the number possessing the characteristic, and follow the asterisks (if any) across to the first figure encountered. This figure is the approximate coefficient of variation.

Rule 2: **Estimate of proportion or percentage possessing a characteristic**

An estimate of the proportion or percentage in a domain possessing a characteristic is formed as the ratio of the estimate of the number in the domain possessing the characteristic (numerator) and the estimate of the number of individuals in the domain (denominator). The approximate coefficient of variation depends on both the ratio and the numerator, except if the domain coincides with the level of estimation covered by the table. In this case the approximate CV of the estimate of the proportion possessing the characteristic is identical to the approximate CV of the estimate of the number possessing the characteristic, and Rule 1 can be used.

If the domain is a subset of the level of estimation covered by the table, the approximate CV is found using both the estimate of the proportion or percentage (across the top of the table) and the numerator of estimate of the proportion or percentage (down the left side of the table). The intersection of the appropriate row and column gives the approximate coefficient of variation. In addition, in such cases estimates of proportions or percentages possessing a characteristic are relatively more reliable than the corresponding estimate of the number possessing the characteristic. For example, an estimate of the percentage of 22-year-olds in Ontario who were 1995 graduates is more reliable than the estimated number of 22-year-olds in Ontario who were 1995 graduates. (Note that in the tables, the CVs decline in value reading from left to right.)

Rule 3: **Estimate of difference between aggregates, proportions 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, standard error of a difference $(\hat{d}=\hat{X}_1-\hat{X}_2)$ is:

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

where $\hat{X}_1$ is estimate 1, $\hat{X}_2$ is estimate 2, and α_1 and α_2 are the CV of $\hat{X}_1$ and $\hat{X}_2$ respectively.

The coefficient of variation of $\hat{d}$ is given by:

$$\alpha_{\hat{d}} = \sigma_{\hat{d}} / \hat{d}$$

This formula is a good approximation for the CV of a difference between separate and uncorrelated characteristics, but may not be accurate otherwise as it will tend to overstate the CV if $\hat{X}_1$ and $\hat{X}_2$ are positively correlated and understate the CV if $\hat{X}_1$ and $\hat{X}_2$ are negatively correlated.

Rule 4: **Estimate of ratio of aggregates, proportions or percentages**

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 1995 graduates and the numerator is the number of 1995 graduates who indicated that they regretted leaving school.

In the case where the numerator is not a subset of the denominator, (for example, the ratio of the number of female 1995 graduates to the number of male 1995 graduates) 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 the ratio itself. That is, the standard error of such a ratio $(\hat{R} = \hat{X}_1 / \hat{X}_2)$ 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:

$$\alpha_{\hat{R}} = \sigma_{\hat{R}} / \hat{R}$$

so it follows that:

$$\alpha_{\hat{R}} = \sqrt{\alpha_1^2 + \alpha_2^2}$$

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: **Estimate of difference 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.

11.1.1 Examples using the ASVT for Categorical Estimates

The following are examples using actual SLFS data to illustrate application of the first four rules presented above (Rule 5 is merely a combination of Rule 3 and Rule 4). The data concern training (other than high school) taken by respondents (Q35A) and to what types of degrees, certificates or diplomas that training was towards (Q36). The domain of estimation for these examples is individuals in the SLFS survey population whose Province of study in 1995 was Ontario.

Example 1: Estimate of number possessing a characteristic (Rule 1)

Using SLFS data, it is estimated that 326,956 individuals in Ontario had taken some form of training. How does the user determine the coefficient of variation of this estimate?

(1) Refer to the ASVT for Ontario, the level of estimation, in Appendix D.

(2) The estimated aggregate before rounding, 326,956, 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 300,000.

(3) The coefficient of variation for an estimated aggregate is found by referring to the first non-asterisk entry on that row, namely 1.5%.

(4) Therefore the approximate coefficient of variation of the estimate is 1.5%.

Note that the same estimated coefficient of variation would apply to the estimate of the proportion of individuals in Ontario who took some form of training, as in this case the denominator of the proportion is the entire province. The estimate of the proportion would be 326,956/418,085=0.782 or 78.2%.

Example 2: Estimate of proportion possessing a characteristic (Rule 2)

Using SLFS data it is estimated that 41.3% (134,950/326,956) of individuals in Ontario who took training took some training at a college. How does the user determine the coefficient of variation of this estimate?

(1) Refer to the ASVT for Ontario in Appendix D.

(2) Because the estimate is a percentage which is based on a subset of a province (i.e. only on individuals who took training), it is necessary to use both the percentage (41.3%) and the numerator portion of the percentage (134,950) in determining the CV.

(3) The numerator, 134,950, 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 125,000. Similarly, the percentage estimate does not appear as any of the column headings, so it is necessary to use the figure closest to it, namely 40.0%.

(4) The figure at the intersection of the row and column used, namely 5.8%, is the coefficient of variation to be used.

(5) Therefore the approximate coefficient of variation of the estimate is 5.8%.

Note that the approximate coefficient of variation for an estimate of the number of individuals in Ontario who took training at a college would be 6.2%.

Example 3: Estimate of difference between proportions (Rule 3)

Using SLFS data it is estimated that 80.3% (166,305/207,228) of females had taken some form of training, while 76.2% (160,651/210,857) of males had taken some form of training. How does the user determine the coefficient of variation of the difference between these two estimates?

(1) Using the ASVT for Ontario, the approximate CV of the estimate of the proportion of individuals who took

training is 2.1% for females and is 3.7% for males.

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

$$s_{\hat{d}} = \sqrt{(\hat{X}_1 a_1)^2 + (\hat{X}_2 a_2)^2}$$

where $\hat{X}_1$ is estimate 1 (the proportion of females who took training), $\hat{X}_2$ is estimate 2 (the proportion of males who took training), and a_1 and a_2 are the CV of $\hat{X}_1$ and $\hat{X}_2$, respectively.

(3) Inserting the values of our example, $\hat{d}$ = (0.803-0.762)=0.041 and the standard error of the estimate $\hat{d}$ is:

$$s_{\hat{d}} = \sqrt{[(.803)(.021)]^2 + [(.762)(.037)]^2} = 0.0329$$

(4) The approximate coefficient of variation of $\hat{d}$ is given by $s_{\hat{d}} / \hat{d}$ =0.0329/0.041=0.802 or 80.2%. This estimate is "Not for release" since the approximate CV exceeds 33.3%.

Example 4: Estimate of ratio of aggregates (Rule 4)

Suppose now a user wants to compare the number of females in Ontario who did not take training to the number of males in Ontario who did not take training. The user is interested in comparing these estimates in the form of a ratio. How does the user determine the coefficient of variation of this estimate?

(1) First, this estimate is a ratio of which the numerator (= $\hat{X}_1$) is the number of females in Ontario who did not take training and the denominator (= $\hat{X}_2$) is the number of males in Ontario who did not take training.

(2) Refer again to the ASVT for Ontario.

(3) The numerator of this ratio estimate is 40,925. The approximate CV for this estimate is found to be 12.5%, the first non-asterisk entry in the closest row, namely 40,000.

(4) The denominator of this ratio estimate is 50,206. The approximate CV for this estimate is found to be 10.8%, the first non-asterisk entry in the closest row, namely 50,000.

(5) The approximate coefficient of variation of the ratio estimate is given by Rule 4 as

$$a_{\hat{R}} = \sqrt{a_1^2 + a_2^2}$$

where a_1 and a_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively.

(6) Inserting the values of our example, $\hat{R}$ =40,925/50,206=0.815 and the approximate coefficient of variation of $\hat{R}$ is given by:

$$a_{\hat{R}} = \sqrt{(.125)^2 + (.108)^2} = 0.165$$

Therefore the approximate coefficient of variation of the estimate is 16.5%.

11.2 How to use the ASVT for 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 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 ($\hat{X}-k$, $\hat{X}+k$), one below the estimate and one above the estimate, 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 approximate coefficient of variation of the estimate $\hat{X}$, and then using the following formula to convert to a confidence interval Cl:

$$CI_x = (\hat{X} - t\,\hat{X}\,a_{\hat{X}},\ \hat{X} + t\,\hat{X}\,a_{\hat{X}})$$

where $a_{\hat{X}}$ is the determined approximate CV 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 = 3 if a 99% confidence interval is desired.

It should be noted that in the above calculation of a confidence interval the approximate CV must be expressed as a proportion rather than as a percentage (e.g. as 0.05 for a CV of 5%).

11.2.1 Example using the ASVT to obtain Confidence Limits

A 95% confidence interval for the estimate of the proportion of individuals in Ontario who took training (from Example 1 in Section 11.1.1) would be calculated as follows:

$\hat{X}$ = 0.782
t = 2
$a_{\hat{X}}$ = 0.015 (corresponding to an approximate CV of 1.5%)

Cl_x = {.782 - (2) (.782) (.015), .782 + (2) (.782) (.015)}
= {.782-.023,.782 +.023}
= {.759, .805}

With 95% confidence it can be said that between 75.9% and 80.5% of individuals in Ontario did not take training other than high school.

11.3 How to obtain Coefficients of Variation for Quantitative Estimates

For quantitative estimates, special tables would have to be produced to determine their sampling error. Since the variables for the School Leavers Follow-up Survey 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 categorical estimate (i.e., the estimate of the number of persons contributing to the quantitative estimates). If the corresponding categorical estimate is not releasable, the quantitative estimate will not be either. For example, the coefficient of variation of total number of hours worked in a week would be greater than the coefficient of variation of the corresponding proportion of those who actually worked. Hence if the coefficient of variation of the proportion is not releasable, then the coefficient of variation of the corresponding quantitative estimate will not be releasable either.

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 data file into random 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.

References

Corporate-Higher Education Forum. **Learning Goals for K-12 Education: To Be Our Best** Montreal: 1992

Economic Council of Canada. **A Lot to Learn: Education Training in Canada** Ottawa: 1992

Frank J. **After High School The First Years.** The first report of the School Leavers Follow-up Survey, 1995 A report prepared under contract to HRDC and Statistics Canada, Ottawa: Autumn 1996

Gilbert, S.et al. **Leaving School**. Ottawa: Supply and Services Canada, September 1993.

Jones S. **"Relationships among Proxy and Direct Assessments of Basic Skills in the School Leavers Follow-up Survey".** A report prepared under contract to HRDC and Statistics Canada, Ottawa: April 1997.

Jones S. "**The Definition of Basic Skills and Development of Measurement Instruments**" A report prepared as part of a contract with Occupational Systems and Standards, Occupational and Career Information Branch, Human Resources and Labour Canada. Ottawa: November 1993.

Jones S. "Literacy Programming and the Survey of Literacy Skills Used in Daily Activities" in **Adult Literacy in Canada: Results of a National Study** Ottawa: Statistics Canada, Catalogue 89-525, September 1991.

McLaughlin, M. "**Employability Skills Profile: What are Employers Looking For?** Ottawa: The Conference Board of Canada, 1992.

Niece, D., Adsett, M. And Rodney, W. "**Direct Versus Proxy Measures of Adult Functional Literacy: A Preliminary Re-examination**" Working Paper #2, prepared for the joint OECD/UNESCO Seminar in Hamburg, Nov.19-22, 1990. Ottawa: Social Trends Analysis Directorate, Department of the Secretary of State, 1990.

Social Survey Methods Division. **School Leavers Survey Evaluation Report** - Appendix B: Methodological Overview of the 1991 School Leavers Survey, p.B-25, Statistics Canada