2. BACKGROUND

The probable lack of success in the labour market of individuals who leave school before they have received a high school diploma or certificate can result in costs to society as a whole. Poor labour market experience, from the individual's point of view, may result in poor earning potential, poor self-image and a low standard of living. From a socio-economic point of view, the cost can take the form of the loss of potential contribution, higher unemployment insurance payments and/or greater welfare costs.

There is no doubt that level and type of education imparts to the individual the ability to make career choices, thereby affecting their resultant earning capacity and position in society. Of particular interest is the impact of low education since one of the primary focuses of the survey is on those who did not complete secondary school. Comparisons are able to be made between dropouts and those who have completed high school and moved into the labour market and with those who are continuing their education.

3. OBJECTIVES

The primary purpose of the survey was to establish a rate for school leavers for Canada and the Provinces, as well as develop comparative profiles of three groups of secondary school attendees: those who successfully completed school, those still in the education system and those who left 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 a truncated education. As a result of 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:

Transitions

- At what level 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?

Labour Market

- 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?

4. CONCEPTS AND DEFINITIONS

The following section outlines the concepts and definitions used in the “School Leavers Survey”. Users are referred to Appendix A of this document for a copy of the survey form used in the 1991 survey. This document is simply a print-out of the Computer Assisted Telephone Interview (CATI) program 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.

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.

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.

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.

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 non-completers.

Field of Study

Codes up to and including 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

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

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.

Graduates

Respondents who have graduated from high school (Grades 11, 12 or 13 depending on the province). 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.

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 continuers) 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 or looking for work. Full-time students looking for full-time work were also classified as not in the labour force.

Non-Completers

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

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.

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.

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 mentally handicapped, deaf, blind and disabled people.

Standard Occupational Classification (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.

Temporary/Seasonal Job

A job where there is a definite sign that the job will end at some specific point in time.

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, sponsored by Employment and Immigration Canada, which combine high school study with on-the-job training.

5. SURVEY DESIGN

Data for the School Leavers Survey (SLS) was collected between April and June 1991.

5.1 Population Coverage

The survey population base was the Family Allowance (FA) file, which is the best available list of young people in Canada. Since the target population was young adults aged 18 to 20, several years of the FA file were used (1986 to 1990), to come up with the complete population. The FA file also contained potential dropout indicators such as “child has left home”, “child is working”, etc.

Inclusions:

Included in the survey population were people who were 18 to 20 years old as of April 1st, 1991, residing in one of the 10 provinces in Canada.

Exclusions:

Excluded from the survey, before the data collection process began, were potential respondents who 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., an orphanage, a prison, transition homes, shelters etc.).

Because the sample frame was based on the Family Allowance File, it was necessary to exclude what looked like potential respondents from the survey during the data collection process, for example, bank accounts. Because of the instance of large groups of children being cared for at one location, such as a Hutterite Community, it was decided to exclude the third or higher selected child to be interviewed at the same contact point.

5.2 Sample Design

The survey is based on a stratified sample design. The population was stratified by province, age and potential dropout indicator. The potential dropout indicator was derived using codes from the Family Allowance file, and could take 2 states: potential dropout and non-potential dropout.

POPULATION COUNT

Province	Age 18	Age 18	Age 19	Age 19	Age 20	Age 20
	Potential	Non-potential	Potential	Non-potential	Potential	Non-potential
Newfoundland	111	11424	172	11910	181	11703
P.E.I.	25	2003	32	2195	54	2122
Nova Scotia	132	13576	163	14167	175	14357
New Brunswick	185	11998	270	12261	302	12281
Quebec	557	86514	1096	88579	1046	91928
Ontario	1618	135231	1922	137478	2006	142403
Manitoba	144	16174	200	16413	215	16939
Saskatchewan	277	15283	329	15392	356	15563
Alberta	360	34801	499	35567	644	36919
British Columbia	443	40349	460	41159	559	44056

The sample was allocated such that the national and provincial dropout rates could be evaluated within a certain precision. The group of 20 year olds was chosen to estimate that proportion, since they are the oldest, thus having a lower chance of being a continuer in high school since “continuer” is a “temporary state” which results in either a state of being a dropout or graduate. The final age range of 18 to 20 years old inclusive was used so that there was the possibility to, first, evaluate the dropout rates and, second, to be able to compare the three groups of school attendees, namely dropouts, graduates and continuers.

5.3 Sample Size

Based upon the specifications above, a simple random sample was drawn within each strata (sometimes all of units in the strata were sampled), such that the basic objectives (provincial and national dropout rates and comparative profiles of graduates, continuers and dropouts) could be met. A final sample size of 18,000 was decided upon, and is distributed among strata in the following manner:

SAMPLE COUNT

Province	Age 18	Age 18	Age 19	Age 19	Age 20	Age 20
	Potential	Non-potential	Potential	Non-potential	Potential	Non-potential
Newfoundland	111	209	172	173	181	469
P.E.I.	25	105	32	108	54	566
Nova Scotia	132	228	163	217	175	475
New Brunswick	185	165	270	105	302	368
Quebec	557	353	575	415	330	370
Ontario	865	480	742	523	330	370
Manitoba	144	241	200	205	215	435
Saskatchewan	277	163	253	177	356	344
Alberta	360	290	380	270	330	370
British Columbia	443	247	411	289	330	370

5.4 Obtaining Telephone Numbers

The School Leavers Survey was conducted over the telephone. However, the Family Allowance File did not contain telephone numbers. Various methods were used to assign telephone numbers to surname, address and postal code combinations (listed below) to those selected from the file. Through this and other tracing techniques, interviewers were provided with a number of potential telephone numbers to use in tracing selected individuals.

1. match by Surname, exact address, postal code
2. match by Surname, partial address, postal code
3. match by Surname, postal code
4. match by exact address, postal code
5. match by exact previous address, postal code

Thus, each respondent could have several phone numbers that have matched. Some of them would have none.

6. DATA COLLECTION

Data collection for the School Leavers Survey was carried out from April to June 1991. Due to different considerations, such as budget, difficulty and length of the questionnaire, it was decided to do a telephone survey using a Computer Assisted Telephone Interview (C.A.T.I).

6.1 Interviewing for the SLS

The data was centrally collected at Statistics Canada's head office by approximately 40 interviewers, who were Statistics Canada employees, people from agencies and students who were trained specifically to carry out the SLS. They tried to contact each of the sampled persons to obtain the required information; no "proxy" information was collected.

6.2 Supervision and Control

Quality monitoring was performed during the data collection period. The monitoring was performed by listening to the telephone conversation while at the same time, looking at the data that was entered on the computer. Consequently, better quality data were obtained.

6.3 Tracing and Response Rates for the School Leavers Survey

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

	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 10782 youths traced, 9460 youths were complete respondents and in-scope to the survey (88%).

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, 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 Correction of the 1991 School Leavers Survey

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” code was assigned to the item.

7.3 Coding of Open-ended Questions

A few data items on the questionnaire were recorded by interviewers in an open-ended format. 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 and standard occupational 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.

7.5 Weighting

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

The weighting phase is a step which calculates, for each record, what this number is and places it on the microdata file for each record. This weight must be used to derive estimates from the microdata file. For example, if the number of individuals who ever left school is to be estimated, it is done by selecting the records referring to those individuals in the sample with that characteristic and summing the weights entered on those records.

Details of the method used to calculate these sampling weights are presented in Appendix B.

8. SAMPLING ERROR

The SLS 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 is expressed as a percentage of the estimate.

For example, suppose that based upon the survey results, one estimates that 22% of the individuals 20 years of age failed a grade in elementary school and that this estimate is found to have standard error of .006. Then the coefficient of variation of the estimate is calculated as:

$$\frac{.006}{.22} \times 100\% = 2.7\%$$

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 Survey.

(1) Categorical Estimates

Categorical estimates are estimates of the number, portion 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.

In this context, an estimate of the number of persons possessing a certain characteristic is referred to as an estimate of an aggregate.

(2) Quantitative Estimates

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

An example of a quantitative estimate is the average number of 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 the number of individuals who didn't complete primary/secondary school and dropped out more than once.

8.1 Where to Obtain Coefficients of Variation for Category Estimates

In order to supply coefficients of variation which would be applicable to a wide variety of categorical estimates produced from this microdata tape and which could be readily accessed by the user, a set of 'look up' tables, referred to as Approximate Sampling Variability Tables, have been produced and included in Appendix D.

These coefficients of variation are derived using the variance formula for simple random sampling and incorporating a factor which reflects the multi-stage, clustered nature of the sample design. This factor, known as the design effect, 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 which 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.

Estimates of actual variance for specific variables may be obtained from Statistics Canada on a cost recovery basis. As noted in Section 10.3, use of actual variance estimates allows users to release otherwise unreleaseable estimates, i.e. estimates with coefficients of variation in the 'restricted' range.

The following rules should enable the user to determine the approximate coefficients of variation from the Sampling Variability Tables for estimates of the number, proportion or percentage of the surveyed population possessing a certain characteristic and for ratios and differences between such estimates.

Rule 1: **Estimates of Numbers Possessing a Characteristic (Aggregates)**

The coefficient of variation depends only on the size of the estimate itself. On the Sampling Variability Table for the appropriate unit of analysis and the appropriate geographic area, locate the estimated number in the left-most column of the table (headed “Numerator of Percentage”) and follow the asterisks (if any) across to the first figure encountered. This figure is the approximate coefficient of variation.

Rule 2: **Estimates of Proportions or Percentages Possessing a Characteristic**

The coefficient of variation of an estimated proportion or percentage depends on both the size of the proportion or percentage and the size of the total upon which the proportion or percentage is based. Estimated proportions or percentages are relatively more reliable than the corresponding estimates of the numerator of the proportion or percentage, when the proportion or percentage is based upon a sub-group of the population. For example, the percentage of males who ever left school is more reliable than the estimated number of males who ever left school. (Note that in the tables, the cv's decline in value reading from left to right).

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

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

Rule 3: **Estimates of Differences Between Aggregates or Percentages**

The standard error of a difference between two estimates is approximately equal to the square root of the sum of squares of each standard error considered separately. That is, 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 coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively. The coefficients of variation of $\hat{d}$ is given by $\sigma_{\hat{d}}/\hat{d}$. This formula is accurate for the difference between separate and uncorrelated characteristics but is only approximate otherwise.

Rule 4: **Estimates of Ratios**

In the case where the numerator is a subset of the denominator, the ratio should be converted to a percentage and Rule 2 applied. This would apply, for example, to the case where the denominator is the number of individuals who ever left school and the numerator is the number of individuals who ever left school that had a job at the time of the survey.

In the case where the numerator is not a subset of the denominator, (for example, the ratio of the number of individuals who took soft drugs (e.g., marijuana) as compared to the number of individuals who took hard drugs (e.g., crack)) the standard deviation of the ratio of the estimates is approximately equal to the square root of the sum of squares of each coefficient of variation considered separately multiplied by $\hat{R}$. That is, the standard error of a ratio $(\hat{R}=\hat{X}_1/\hat{X}_2)$ is:

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

where α_1 and α_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively.

The coefficient of variation of $\hat{R}$ is given by $\sigma_{\hat{R}}/\hat{R}$. The formula will tend to overstate the error if $\hat{X}_1$ and $\hat{X}_2$ are positively correlated and understate the error if $\hat{X}_1$ and $\hat{X}_2$ are negatively correlated.

Rule 5: **Estimates of Differences of Ratios**

In this case, Rules 3 and 4 are combined. The cv's for the two ratios are first determined using Rule 4, and then the cv of their difference is found using Rule 3.

The following two 'real life' examples are included to assist users in applying the foregoing rules.

Example 1

Suppose that a user estimates that 169,536 youths 18 to 20 in Ontario never played video games in an arcade. How does the user determine the coefficient of variation of this estimate?

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

(2) The estimated aggregate, 169,536, 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 150,000.

(3) The coefficient of variation for an estimated aggregate is found by referring to the first non-asterisk entry on that row, namely, 4.3%

(4) Therefore the approximate coefficient of variation of the estimate is 4.3%

Example 2

Suppose that the user estimates that 76.7% of youths 18 to 20 in Ontario never played video games in an arcade. This is the expression of the estimate obtained in Example 1, the number of youths 18 to 20 in Ontario who never played video games in an arcade, as a percentage of all youths 18 to 20 in Ontario. How does the user determine the coefficient of variation of this estimate?

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

(2) Because the estimate is a percentage which is based on a subset of the total population, it is necessary to use both the percentage (76.7%) and the numerator portion of the percentage (169,536) in determining the coefficient of variation.

(3) The numerator, 169,536, does not appear in the left-hand column (the 'Numerator of Percentage' column) so it is necessary to use the figure closet to it, namely 150,000. 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, 70.0%

(4) The figure at the intersection of the row and column used, namely, 3.0% is the coefficient of variation to be used.

(5) Therefore the approximate coefficient of variation of the estimate is 3.0%.

8.2 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 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 microdata files into subgroups (or replicates) and determining the variation in the estimate from replicate to replicate. Users wishing to derive coefficients of variation for quantitative estimates may contact Statistics Canada for advice on the allocation of records to appropriate replicates and the formulae to be used in these calculations.

8.3 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, one below the estimate and one above the estimate, as ($\hat{X}-k$, $\hat{X}+k$) 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 Sampling Variability Tables by first determining from the appropriate table the coefficient of variation of the estimate $\hat{X}$, and then using the following formula to convert to a confidence interval CI:

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

where $\alpha_{\hat{X}}$ is the determined coefficient of variation of $\hat{X}$

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

Example

A 95% confidence interval for the estimated proportion of youths 18 to 20 year old in Ontario who never played video cames in an arcade (from example 2 above) would be calculated as follows.

X = 76.7% (or expressed as a proportion = .767)

t = 2

$\alpha_{\hat{X}}$ = 3.0% (.030 expressed as a proportion)

CI_x = {.767 - (2) (.767) (.030), .767 + (2) (.767) (.030)}
={.767-.046,.767 +.046}
= {.721, .813}

With 95% confidence it can be said that between 72% and 81% of youths of 18 to 20 year old in Ontario never played a video game in an arcade.

9. 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 errors.

Over a large number of observations, randomly occurring errors will have little effect on estimates derived from the survey. However, errors occurring systematically will contribute to biases in the survey estimates. 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 SLS 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.

9.1 Total Non-response

Total non-response is a major source of non-sampling error in many surveys. In the SLS total non-response occurred because the interviewer was either unable to contact the respondent, or the respondent refused to participate in the survey. Total non-response was handled by adjusting the weight of those who responded to compensate for those who did not respond.

9.2 Partial Non-response

Partial non-response in the SLS 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 SLS.

10. PUBLICATION AND RELEASE GUIDELINES

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 tapes of the School Leavers 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 tapes. With the aid of these guidelines, users of microdata should be able to produce the same figures as those produced by Statistics Canada and, at the same time, will be able to develop currently unpublished figures in a manner consistent with these established guidelines. This section consists basically of four sub-sections - the rounding guidelines, the sample weighting guidelines, the sampling variability guidelines and guidelines for statistical analysis.

10.1 Rounding Guidelines

In order to ensure that estimates for publication or other releases derived from these microdata tapes 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 thousands 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 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 microdata tape records. Otherwise, the estimates derived from the microdata tapes 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.

10.3 Sampling Variability Guidelines

Before releasing and/or publishing any estimate from these microdata tapes, 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. Once the coefficient of variation for this estimate has been determined, the user should follow the guidelines below.

Sampling Variability Guidelines for
The School Leavers Survey

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. Confidential	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 (---).

Note: These sampling variability guidelines should be applied to rounded estimates.

10.4 Guidelines for Statistical Analysis

The SLS is based upon a design with stratification, and unequal probabilities of selection of respondents. Using data from such surveys presents problems to analysts because the survey design and the selection probabilities affect the estimation and variance calculation procedures that should 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.