

Youth in Transition Survey (YITS)
18 –20 year-old cohort
Cycle 1

USER GUIDE

May 30, 2003

Statistics Canada Statistique Canada

1.0 Introduction .....

2.0 Background 6

3.0 Objectives 7

4.0 Concepts and Definitions 8
4.1 YITS Concepts 10
4.1.1 Education – high school, junior high or elementary 8
4.1.2 Post-secondary Education 13
4.1.3 Education and work aspirations (Module M) 16
4.1.4 Health – Activity limitations (Module M) 16
4.1.5 Training for a job or career (Module N) 16
4.1.6 Employment (Modules P1, P2 and P5) 16
4.1.7 GAPS - months not in school full-time and not working (Module PS) 20
4.1.8 Volunteering (Module Q) 21
4.1.9 Skills (Module R) 21
4.1.10 Income (Module U) 21
4.1.11 Personal Characteristics and Family Background 22
4.2 Derived Variables 25
4.2.1 Elementary and high school education 27
4.2.2 Post secondary education 30
4.2.3 Financing post secondary education 35
4.2.4 Employment 36
4.2.5 GAPS - Months not in school full-time and not working 39
4.2.6 Activities over each month in 1999 41
4.2.7 Volunteer 41
4.2.8 Income 42
4.2.9 Demographics and family background 42
4.3 YITS Scales 45
4.3.1 Defining scales and their uses 45
4.3.2 Scale Development 46
4.3.2.1 Investigation of model validity 47
4.3.2.2 Estimation of the scores 48
4.3.2.3 Scale reliability 49
4.3.2.4 Scale validity 50
4.3.3 The high school engagement scales 50
4.3.3.1 Why is there interest in a school engagement model? 50
4.3.3.2 What is the YITS high school engagement scale? 51
4.3.3.3 Structural model of high school engagement 52
4.3.4 Analysis of high school engagement scale 52
4.3.4.1 Testing model validity 52
4.3.4.2 Estimating Scores 53
4.3.4.3 Testing scale reliability 54
4.3.4.4 Testing scale validity 54
4.3.5 Scale references 56

5.0 Survey Methodology 57
5.1 LFS Sample Design 57
5.2 LFS Target population 57
5.3 LFS Stratification 57
5.4 LFS Cluster delineation and selection 58
5.5 LFS Dwelling selection 59
5.6 LFS Sample rotation 59
5.7 Household members eligible for the LFS 59
5.8 YITS Target population and sample design 59

5.9 YITS Sample Size By Province 61

6.0 Data Collection 62
6.1 The Questionnaire 62
6.2 Training 62
6.3 Supervision and Control 62
6.4 The Interview 62
6.5 Follow-up of Non-response 63

7.0 Data Processing. 63
7.1 Data capture 63
7.2 CAI Editing 63
7.3 Minimum completion requirements 64
7.4 Head Office editing 64
7.5 Creation of Derived Variables 65
7.6 Imputation of missing data for income and earnings variables 66
7.7 Weighting 67

8.0 Data Quality 68
8.1 The Frame 68
8.2 Measuring sampling error 69
8.3 Non-sampling error 71
8.4 Response rates 72

9.0 Guidelines for Tabulation, Analysis and Release 74
9.1 Rounding Guidelines 74
9.2 Sample Weighting Guidelines for Tabulation 75
9.2.1 Definitions of types of estimates: Categorical vs. Quantitative 75
9.2.2 Tabulation of Categorical Estimates 76
9.2.3 Tabulation of Quantitative Estimates 76
9.3 Guidelines for Statistical Analysis 76
9.4 CV Release Guidelines 77

10.0 Weighting 79
10.1 Weighting procedures for the LFS 79
10.1.1 LFS Basic weight 79
10.1.2 LFS Cluster sub-weight 79
10.1.3 LFS Stabilisation weight 79
10.1.4 LFS Non-response adjustment 80
10.1.5 LFS sub-weight 80
10.1.6 LFS Calibration 80
10.2 Weighting procedures for YITS 80
10.2.1 YITS Adjustment for 36 groups 81
10.2.2 YITS Household non-response 81
10.2.3 YITS Sampling within household 82
10.2.4 YITS Non-response of selected person 82
10.2.5 YITS Post-stratification 82

11.0 Variance estimation 83

1.0 Introduction

The Youth in Transition Survey (YITS) is a longitudinal survey designed to provide policy-relevant information about school-work transitions and factors influencing pathways among education, training and work. The first data collection – Cycle 1of the survey - for the cohort aged 18 to 20, was conducted by Statistics Canada between January and April 2000 with the co-operation and support of Human Resources Development Canada.

This User guide for cycle 1 of the YITS 18 to 20-year-olds is developed for the 2nd release of the microdata file. The first release was in May 2002.

It should be pointed out that not all the data collected in the first cycle of the YITS 18 to 20-year-old were included in the first microdata file. What was included in that release are those data that are discussed in the analytical report produced by Human Resources Development Canada and Statistics Canada: "At a Crossroads: Results for the 18-20-Year-Old Cohort of the Youth in Transition Survey 2000".

The second release, as of March 2003, includes the complete microdata set with derived variables and weights.

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

At Statistics Canada:

Client Services
Centre for Education Statistics, Statistics Canada
2000 Main Building
Tunney's Pasture
Ottawa, Ontario K1A 0T6
Telephone: (613) 951-7608
Toll free: 1-800-307-3382
Fax: (613) 951-9040
E-mail: educationstats@statcan.ca

At Human Resources Development Canada:

Urvashi Dhawan-Biswal, Survey Co-ordinator
Child, Youth and Social Development Studies
Applied Research Branch, Human Resources Development Canada
165 Hotel de Ville Street, Phase II, 7th Floor
Hull, Quebec K1A 0J2

2.0 Background

Starting in 1996, Human Resources Development Canada and Statistics Canada began developing the Youth in Transition Survey (YITS). Consultations took place with representatives from federal government departments with an interest in youth policy, provincial ministries and departments of education and labour, practitioners working directly with youth (teachers, counsellors, school board personnel and social workers), employers, business and education associations, academic researchers, youth and parents[1]. The result of these consultations was the development of the YITS as a longitudinal survey to collect policy-relevant information on the school-work transitions of young people, and the factors that influence such transitions.

The Youth in Transition Survey is composed of a family of surveys. These surveys provide a set of information from which complex data analysis between each survey can be done. One cohort was students who were 15 year-olds at the time of data collection. The Canadian YITS was first administered alongside the Program for International Student Assessment (PISA) in 2000. This cohort is referred to as the "PISA/YITS Reading Cohort". In 2003 another survey will be conducted and is referred to as the "PISA/YITS Mathematics Cohort".

The second cohort interviews were conducted with 18 to 20 year-olds who were selected to provide information on the post-secondary and labour market experiences of young adults. These youth are to be re-interviewed every two years until young adulthood to continue to obtain information at different stages in their lives. In future this cohort will be referred to as the "18-20 year-old YITS Cohort".

This guide refers solely to the survey of the 18 to 20-year-old YITS cohort.

[1]For more information about the consultation process and other aspects of YITS, see *Youth in Transition Survey Project Overview – T-00-5E (September 2000)* (Ottawa: Human Resources Development Canada, 2000, Cat. No. MP32-30/00-5E/F)

3.0 Objectives

The broad objectives of the Youth in Transition Survey are:

- to examine key transitions in the lives of youth, such as the transition from high school to post-secondary schooling and the initial transition from schooling to the labour market;
- to better understand educational and labour market pathways and the factors influencing these pathways;
- to identify educational and occupational pathways that provide a smoother transition to the labour market;
- to examine the incidence, characteristics, factors and effects of leaving school;
- to understand the impact of school effects on educational and occupational outcomes;
- to examine the contribution of work experience programs, part-time jobs, and volunteer activities to skill development and transition to the labour market;
- to study the attitudes, behaviours, and skills of young people entering the labour market;
- to gain a better understanding of the determinants of post-secondary entry and post-secondary retention, including education financing;
- to better understand the role of educational and labour market aspirations and expectations in investment in further education and career choice; and,
- to explore the educational and occupational pathways of various sub-groups, particularly youth “at risk”.

4.0 Concepts and Definitions

Major data elements

This section outlines the major concepts and definitions of interest to the users of the YITS microdata file. The reference period for the data collected in the YITS is up to December 1999. The only exceptions are the personal characteristics and family background data, because respondents were asked for their current situation; in other words, as of the date of the interview.

Section 4.1 describes the concepts included in the YITS and section 4.2 describes derived variables included on the data file.

4.1 YITS Concepts

There are many concepts that are pertinent to understanding the YITS questionnaire. They have been categorised under different questionnaire modules pertaining to education, health, training, employment, volunteer, skills, income, and personal characteristics and family background.

Sub-Section	Description	Modules
4.1.1	High school, junior high or elementary education School courses School – attitudes and behaviours	B C F
4.1.2	Post-secondary education School – attitudes and behaviours School – financing	H K L
4.1.3	Education and work aspirations	M
4.1.4	Health – activity limitations	M
4.1.5	Training for a job or career	N
4.1.6	Employment	P1 and P2
4.1.7	GAPS - months not in school full-time and not working	PS
4.1.8	Volunteering	Q
4.1.9	Skills	R
4.1.10	Income	U
4.1.11	Personal characteristics and family background	V

4.1.1 Education – high school, junior high or elementary

The high school education section of the YITS contains questions on the respondent's educational attainment, experiences, and course levels and grades in language arts and math.

This section first asks the respondent's current enrolment status. Youth who report that they are not enrolled provide their reason for leaving school and the date at which this separation occurred. The respondents also state the highest grade level they have taken, the highest grade level they have completed, and whether they met the requirements for high school graduation. In addition, the survey asks all youth if they ever dropped-out of high school. Those who dropped-out of high school state the number of times they have done so.

Adult high school (Module B): education taken by adults to obtain their high school diploma or equivalencies, and can be taken in an alternative format or non-traditional setting (internet, at malls, television registration, correspondence or distance learning).

Alternate programs (Module B): high school programs flexible enough to accommodate students who, otherwise might dropout, students returning to school, pregnant teens, and

adults. Students study at their own pace and sometimes study at non-traditional school sites, such as shopping malls or office buildings.

Alternative school (Module B): a school offering a provincially approved curriculum that uses different teaching methods or places, with the emphasis on teaching cultural identity. Parents frequently work with the teachers in the classroom and in planning programs.

Continuers – high school (Module B): respondents who were continuing their studies at elementary, junior high or high school in December 1999. Respondents continuing with post-secondary are not classified as high school continuers.

Correspondence courses (Module B): an educational or training activity that does not require students to physically attend a school, college or university. Mail, radio and television or other media communications such as the Internet are methods used to deliver the instruction. These courses are recognised by the province and exams are conducted in accordance with provincial standards.

Education: Elementary / Junior high / High school (Modules B and C): the responsibility for education in Canada rests with provincial and territorial governments. Each province and territory has developed its own system for education, and the structure can differ from jurisdiction to jurisdiction. The following table illustrates the similarities and differences for most schools, as of 1998-99.

Province/ Territory	**Pre-grade 1 (by grade)**	**Elementary (by grade)**	**Junior High/ Intermediate/ Middle school (by grade)**	**Senior High/ High school (by grade)**
Newfoundland and Labrador	5 yr-old kindergarten	1-6	7-9	Levels I-III (10-12)
Prince Edward Island		1-6	7-9	10-12
Nova Scotia	Pre-grade 1	1-6	7-9	10-12
New Brunswick (English Sector)	5 yr-old kindergarten	1-5	6-8	9-12
New Brunswick (French Sector)	5 yr-old kindergarten	1-8		9-12
Quebec	4 & 5 yr-old kindergarten	1-6		Secondary 1-5
Ontario	4 & 5 yr-old kindergarten	1-8		9-12 and OAC/ (grade 13)*
Manitoba	4 & 5 yr-old kindergarten	1-8		Senior 1-4 (9-12)
Saskatchewan	5 yr-old kindergarten	1-5	6-9	10-12
Alberta	5 yr-old kindergarten	1-6	7-9	10-12
British Columbia	5 yr-old kindergarten	1-7		8-12
Yukon	5 yr-old kindergarten	1-6	7-9	10-12
Northwest Territories	5 yr-old kindergarten	1-6	7-9	10-12

* OAC/(grade13) is being phased out, and will no longer exist beginning in September 2003

Elementary school (Module B): the educational structure varies across the provinces (see definition for ***Education: Elementary / Junior high / High school***). The elementary school level is the first level of instruction of children in the current school system. In general, at the elementary grade level, education is general and basic, and as a minimum includes grades kindergarten through six.

Ever dropped out (Module B): question is asked of: high school graduates who at some point dropped out of school, but returned to continue their education until graduation; high school continuers who at some point dropped-out of school, but returned to continue their education; and school leavers that are those who have never graduated.

Full-time/part-time school status (Module B and H): full-time/part-time status is determined by the educational 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 institution attended.

Full-time schooling (Module B): full-time schooling is schooling or courses taken as a full-time student (see ***Full-time/part-time studies***).

Graduates – high school (Module B): respondents who have completed the minimum requirements for a high school graduation certificate, diploma or equivalent are considered to have graduated (see definition for ***High school graduation***). Some people might still take courses even after they have obtained their graduation certificate, for reasons such as upgrading marks, or taking courses not taken previously.

High school (Module B): the educational structure varies across the provinces (see definition for ***Education: Elementary / Junior high / High school***). In general, at the high school level there is usually a choice of at least two programs: academic or vocational. Some secondary schools may specialise in vocational training (technical and commercial) but most high schools offer both academic courses (preparatory to university) and vocational courses, which prepare students either for an occupation or for further post-secondary non-university education.

High school graduation - diplomas and equivalencies (Module B): the following table lists the graduation diploma, certificate or equivalency awarded by province:

Province	Graduation diploma, certificate or equivalent
Newfoundland and Labrador	➵Senior High Graduation Diploma ➵Provincial High School Graduation Certificate ➵High School Equivalency Diploma ➵General Educational Development (GED)
Prince Edward Island	➵High School Graduation Certificate ➵General Educational Development (GED)
New Brunswick (English and French Sectors)	➵High School Diploma or Certificate ➵Adult High School Diploma ➵General Educational Development (GED)
Nova Scotia	➵High School Graduation Diploma ➵N.S. Secondary School Equivalency Certificate ➵General Educational Development (GED)
Quebec	➵Diplôme d'Études secondaires (DES)/Secondary School Diploma (SSD, general education) ➵Diplôme d'Études professionelles (DEP)/Secondary School Vocational Diploma (SSVD) ➵Attestation de Formation professionelle (AFP)/Secondary School Vocational Certificate (SSVC) ➵Attestation de Spécialisation professionnelle (ASP)/Attestation of Vocational Studies (AVS)/Attestations d'Équivalence du niveau de scolarité (AENS/AESS)
Ontario	➵Ontario Secondary School Diploma (OSSD) ➵General Educational Development (GED)

Province	Graduation diploma, certificate or equivalent
Manitoba	➵Provincial Diploma ➵Adult Basic Education Certificate ➵General Educational Development (GED)
Saskatchewan	➵Record of Secondary Level standing ➵High School Equivalency Diploma ➵General Educational Development (GED)
Alberta	➵Alberta High School Diploma ➵Certificate of Achievement ➵High School Equivalency Diploma for Mature Students ➵General Educational Development (GED)
British Columbia	➵Certificate of Graduation or "Dogwood Diploma" ➵School Completion Certificate ➵General Educational Development (GED)
Northwest Territory	➵Northwest Territories Secondary School Graduation Diploma ➵General Secondary School Graduation Certificate (phased out as of August 31, 1998) ➵Advanced Secondary School Graduation Certificate (phased out as of August 31, 1998)
Yukon	➵Yukon Graduation Certificate ➵School Completion Certificate

Junior high / Intermediate / Middle school (Module B): the educational structure varies across the provinces. (See definition for ***Education: Elementary / Junior high / High school***). A school forming a link between elementary and secondary education usually consists of grades 7 to 9, which is not common to all provinces.

Leavers / Non-completers - high school *(Module B)*: respondents who had not completed the high school graduation requirements, and were not attending elementary, junior high or high school in December 1999.

Private Elementary/ Secondary (high) school (Module B): in contrast to public schools, private schools are not publicly supported but receive funding from private individuals or groups (e.g., student tuition fees, religious groups). These schools, whether church-affiliated or non-sectarian, are operated and administered by private individuals or groups.

Secondary School Vocational Diploma (SSVD/DEP) (Module B): Although this was a category for the highest level of education completed in Module B, only a few respondents who had taken schooling in Quebec reported this. The information was reported and captured in Module H as to the type of program in which a respondent was working towards, or had received, a degree, diploma or certificate. SSVD can take from 6 months to two years to complete. Requirements to begin SSVD are completion of Secondary III, although some exceptions may be granted. SSVD prepares people for employment in specific industries, such as Buildings and Public works, Motorized Equipment Maintenance; Forestry and Pulp and paper, Health Services and Beauty Care.

Work experience program(s) (Module B): programs or courses, combined with high school which provide students with hands-on experience while spending time with an employer, outside the classroom environment. They are part of the student's curriculum. The student receives credit for participation, and may or may not be paid for the work they do. These programs do not include field trips.

Work experience/preparation programs have various names by province such as school-to-work program, Co-op education, entrepreneurship education, youth apprenticeship, bridges transition-to-work programs, practical and applied arts program, work study component, trade program or information technology. Province-specific examples are provided in the questions.

4.1.2 Post-secondary Education

The post-secondary education section of the YITS contains questions on the respondent's participation in any education that could be counted towards a degree, certificate or diploma from an educational institution. An eligible post-secondary program is one that is above the high school level; is towards a diploma, certificate or degree; would take someone three months or more to complete; and the respondent should have started the program before January 2000. All respondents, who in December 1999, were no longer in high school, or who had completed the high school graduation requirements were asked if they had taken any post-secondary education prior to January 2000. The order of the institutions and programs within an institution were provided by the respondent. Respondents were asked to identify institutions and programs beginning with the most recent. Researchers may want to select one institution in particular on which to do analyses, and may want to look at a particular variable such as HLATT_i (i=1,2, 3 or 4 institution) which indicates the status of the respondent at that institution as being a "continuer" or "non-continuer".

Trades programs offered through apprenticeship, vocational schools or private trade schools do not always require high school graduation. Such education is considered as post-secondary.

The section collects information on the type of post-secondary education:
1) Trades certificate or diploma from a vocational or apprenticeship training;
2) Non-university certificate or diploma from a community college, CEGEP, school of nursing, etc.;
3) University certificate below bachelor degree;
4) Bachelor degree; and
5) University degree or certificate above bachelor degree.

The section asks youth, who report being enrolled in a post-secondary program, for the number of institutions attended, and the number of programs taken altogether. Information collected includes the type of degree sought and the duration of the program. For each program, the survey gathers data on primary and secondary fields of study. Survey staff used this information to derive the Major Field of Study (MFS)[2] code. MFS codes are available on the data file. The youth also report on attitudes and behaviours while taking post-secondary education.

Finally, everyone who had taken a post-secondary program was asked about their sources and the amounts of income used to finance their post-secondary education.

Apprenticeship Programs (Module H): lead to journeyman status in several designated trades. Skills and knowledge are provided through on-the-job experience (components) with short periods of formal instruction. Some examples of apprenticeship trades are auto mechanic, hairdresser, boilermaker, steamfitter, millwright, electrician, plumber, machinist, chef.
On-the-job training and the formal schooling are all counted as part of the formal education in order to obtain a license in a trade.

Bursary (Module L): refers to monetary award to assist a student in the pursuit of his/her studies based on financial need and satisfactory achievement.

[2] See section 4.2.2

College Post-Diploma or Graduate Program (Modules H and M): is a relatively new type of program offered by some colleges. Students usually require a previous college diploma or university degree for admission. Examples of these types of programs are "a certificate in telecommunication management"; "a certificate in international business administration".

Commercial school (Module H): private schools that receive no public funding and are licensed by a province. They engage in providing professional and vocational training for profit.

Community Colleges (Module H): includes 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, or teachers' college and establishments providing technological training in specialised 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 – post-secondary (Module H): respondents who were continuing their studies towards the completion of a post-secondary program in December 1999.

Degrees - first professional (Modules H and M): a first professional degree may be taken part way through or after a university bachelor's degree. Examples of this type of degree are medicine, dentistry, veterinary medicine, law, optometry and divinity. Engineering is not considered to be a first professional degree.

Degrees versus Diplomas (Module H): are different types of PROGRAMS, but the word diploma is sometimes used (incorrectly in English) to refer to either a degree or diploma. Most degrees (but not all) are for a program of study at a university. If the official name of the qualification contains the word "degree", "Bachelor's, Master's or Doctor of", it is a degree. Diplomas are less common from a university, but more common from other institutions such as colleges.

Distance Education or Correspondence program (Module H): an educational or training activity that does not require students to physically attend a school, college or university. Mail, radio and television or other media communications such as the Internet are methods used to deliver the instruction. These courses are recognised by the province and exams are conducted in accordance with provincial standards.

Eligibility, program (Module H): to be deemed eligible, a post-secondary program must meet the following criteria: the program is above the high school level; the program is towards a diploma, certificate or degree; the program would take someone three months or more to complete; and the respondent started taking the program before January 2000. If at least one program within a given institution has been deemed eligible, then the institution itself is deemed eligible.

- Eligible programs include: diplomas, degrees, certificates or licenses obtained through professional associations such as in accounting, banking, real estate or insurance.
- Ineligible programs include: non-professional health certificates such as St. John's First Aid, Red Cross; continuing education or personal interest courses.

Ever dropped out (Module H): question is asked of: post-secondary graduates who at some point dropped out of their program, but returned to continue their education until graduation; post-secondary continuers who at some point dropped-out of their program but returned to continue their education; and school leavers that are those who have never graduated.

Full-time/part-time school status (Module H): full-time/part-time status is determined by the educational 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 institution attended.

Full-time schooling (Module H): full-time schooling is schooling or courses taken as a full-time student (see ***Full-time/part-time studies***).

Graduates – post-secondary (Module H): respondents who have completed the graduation requirements towards a diploma, certificate or degree.

Grants (Module L): is a gift (usually a sum of money) made by a government or corporation (as an educational or charitable foundation) to a beneficiary on the condition that certain terms are accepted or certain engagements fulfilled.

Leavers / Non-completers – post-secondary (Module H): respondents who had not completed the graduation requirements for their program, and were no longer taking courses toward the completion of the program in December 1999.

On-the-job experience program (Module H): programs or courses, combined with post-secondary study, which provide students with hands-on experience while spending time with an employer, outside the classroom environment. They are part of the student's curriculum. The student receives credit for participation, and may or may not be paid for the work they do. For respondents who have participated in such a program, the type of program is collected (e.g., Co-op program, Apprenticeship, Trade/vocational, or another type).

Private training institution (Module H): these are privately-owned schools which are profit oriented and are engaged in providing professional and vocational training, and are licensed by the province.

Programs – post-secondary (Module H): includes programs lasting three months or longer and are above the high school level.

- University programs leading to bachelors, master's or doctoral degrees, or specialised certificates or diplomas.
- Programs offered at CEGEPs, community colleges, technical schools, hospital schools of nursing and similar institutions (towards a diploma, certificate or degree) normally requiring secondary school completion or its equivalency for admission.
- Police Academies; RCMP colleges and training camps; Firefighters' training.

Scholarships, award or prizes (Module L): refers to monetary award to assist a student in the pursuit of his/her studies, based usually on outstanding academic achievement rather than on financial need.

Secondary School Vocational Diploma (SSVD) (Module H): SSVD can take from 6 months to two years to complete. Requirements to begin SSVD are completion of Secondary III, although some exceptions may be granted. SSVD prepares people for employment in specific industries, such as Buildings and Public works, Motorized Equipment Maintenance; Forestry and Pulp and paper, Health Services and Beauty Care.

Trade /vocational certificate or diploma (Modules H and M): this term is used to classify skill courses that prepare trainees for occupations not at the professional or semi-professional levels. A trade-vocational program prepares people for employment in a specific occupation such as a heavy equipment operator, automotive mechanic and upholstering. Many community colleges or technical institutes offer certificates or diplomas at the trade level.

University (Modules H and M): an independent institution granting degrees in at least arts and sciences.

Vocational or Trade School (Module H): 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.

4.1.3 Education and work aspirations (Module M)

The survey asks youth the level of education they would like to get, and the level they think they will get. Respondents are then asked if there is anything standing in their way of going as far in school as they would like. The list includes barriers such as financial situation, not being accepted into a program, wanting to stay close to home, caring for children, etc. Respondents are also asked the kind of job or occupation they would be interested in having when they will be about 30 years old.

4.1.4 Health – Activity limitations (Module M)

All respondents provide information about any long term physical condition, mental condition or health problem that limit the kind or amount of activity they can do at school or at work**. Long term** condition(s) were defined as those that have lasted or are expected to **last six months or more**. The purpose of the disability questions is not to determine the nature of the condition so much as the impact on activities, particularly at school and at work.

4.1.5 Training for a job or career (Module N)

All respondents are asked about their participation, during 1999, in training programs related to a job or towards their career. Respondents, who state that they have participated in such programs, are asked follow-up questions about the skills they have learned through the training program(s) (new computer skills, skills to operate other specialised equipment or machinery, skills to help work with others, and skills to help choose or search for a job).

4.1.6 Employment (Modules P1, P2 and P5)

Determining the type of labour market data to be collected by the YITS presented a challenge. For example, to measure school-work transitions, it was necessary to collect the first job at which the respondent worked after leaving full-time schooling. Measuring all jobs since this event was not feasible. However, as most youth aged 18 to 20 are still in full-time schooling or have left not too long ago, measuring work activities in 1999 represented a good alternative. For those who had left full-time schooling prior to 1999, additional questions on the first job after leaving full-time schooling were added in order to capture the "transition" job. And for those who had not worked in 1999 and had not held a job after leaving full-time schooling, questions were asked to determine if they had ever worked and if so, certain details on that job were collected.

Three different categories of labour market activity were collected: employee jobs, self-employment jobs, and jobs at the farm or business of a family member living in the same household.

Youth's labour market experiences were captured for 1999, in addition to the first job after leaving full-time schooling, if this event occurred prior to 1999. Start and stop dates were collected for a maximum of six jobs the respondent worked at in 1999 (Module P1). Further job details were collected on a maximum of 4 jobs the respondent worked at in 1999 (Module P2). These restrictions were imposed to limit the time of interview and minimise respondent burden. Findings from earlier testing of the survey (the YITS 1999 Pilot with youth aged 20-22) indicated that approximately 2% of youth held more than 4 jobs over a one year period.

Bonuses (Module P2): in some situations, wages are paid in the form of both regular pay cheques and periodic bonuses based on work performance. In these cases, the bonuses

should be averaged over the period for which it applies and included with the wages or salary reported. (See ***Wages or Salary***).

Business, for self-employed persons (Module and P2): for self-employed persons, a business exists when one or more of the following conditions are met:

- an office, store, farm or other place of business is maintained and is used exclusively for conducting the business;
- or the enterprise is incorporated (see definition of incorporated business);
- or the self-employed person usually has paid help;
- or land, buildings, machinery or equipment in which the person has invested money is used by respondents or their employees solely in conducting the business.

Examples of self-employed persons **WITH a business** would be:

- a person with their own beauty salon(s);
- a person with a medical practice;
- someone who sub-contracts from someone else.

Examples of self-employed persons **WITHOUT a business** would be:

- a cleaning person working for a number of people in their homes;
- a freelance writer, a tutor, general handyman or a babysitter who regularly works for a number of people.

Dates of jobs:

Start date of job (Module P1): if the respondent...

- works for the same employer on a "seasonal" basis, the date first started work is the date of the most recent period of uninterrupted work, not the date when he/she first began to work for this employer.
- is a paid worker, who works strictly on-call, the date first started work is the date in which the most recent period of work began. Note: Paid on-call workers are only considered to have a job in those months in which some work was done. Any period of one month or more in which no work was done is considered to be a break in employment for on-call workers, and hence, counted as separate jobs.
- is a paid worker who seeks and obtains employment only at certain times of the year (e.g., students who only work in the summer months), the date first started work is the date when he/she last began to work for this employer, even is he/she has worked for the same employer previously (e.g., last three summers).
- is self-employed with a business, the date first started work is the date when he/she created or acquired the business. For self-employed persons with a business, periods of inactivity are not considered as breaks in employment.
- is self-employed without a business, the date first started work is the date in which the most recent period of continuous work began. Self-employed persons without a business are only considered to have a job in those months in which some work was done. Any period of a month or more in which no work was done is considered to be a break in employment, and hence, the next work period becomes a separate job.

End date (Module P1): if the respondent…

- is a paid worker with a definite schedule of work, the date last worked is the month and year the respondent last worked at his/her job prior to January 2000. Respondents who have had paid leaves from their employer, such as vacation, training or sick leave are included as having worked.
- is a paid worker without a definite schedule of work, the date last worked is the month and year the respondent last worked at this job.
- is self-employed with a business, farm or professional practice, the date last worked is the month and year they ceased the operations of their business, or the business closed down, or December 1999 if they still operated the business at that time.
- is self-employed without a business, the date last worked is the month and the year in which they last did any work.

Eligibility, job (Module P1): to be deemed eligible, a job collected for the 1999 reference period must meet the following criteria: if the job began in 1999, the respondent had to be able to provide the job's start and end dates. If, at the time of interview, the respondent was still working at the job, the end date for that job was set to December 1999.

Employee (Module P2): a person who works for others (i.e. works for an employer) and receives a wage or salary. The employer usually deducts and remits from the wage or salary income tax, Canada/Quebec Pension Plan premiums, etc. There are cases where persons receive a wage or salary but no deductions are made for tax or EI/CPP because the wages earned are too low.
(See ***Self-employed*** and ***Unpaid family worker***.)

Employers (Module P2): Are those persons or companies for whom the respondent did any paid jobs whether part-time or full-time.

First Job (Module P1): First job after leaving full-time schooling identifies the job a respondent held at the time of leaving full-time schooling or the job s/he first started after leaving full-time schooling. During survey collection, respondents who were no longer full-time students in December 1999 were asked to report the first job they worked at after leaving full-time schooling. A procedure was then created to validate the job reported by respondents as their first job after leaving full-time schooling and/or when not reported, to identify one of the other jobs reported as being "first job"

Full-time employment (Module P2): consists of persons who usually work 30 hours or more per week at their job.

Hours of work (Module P2) – Usual number of hours worked -

- Number of paid hours usually worked is asked of employees.
- Number of hours usually worked (paid not part of the question) is asked of self-employed workers and unpaid workers in the family farm or business.

For people who do not work a fixed number of hours, usual hours of work mean the average number of hours during a four-week period. In the survey, usual hours of work are collected for two reference periods. The first reference period is when the respondent last worked at his/her job, and the second is when the respondent first worked at his/her job.

For self-employed workers, number of hours worked include time spent on work-related activities in addition to time actually spent on producing goods or providing services. These related activities include: time spent actively looking for potential clients, preparing estimates, quotes or tenders; time spent on operating a business, professional practice or farm even if no sales were made, no professional services were rendered or nothing was actually produced; time spent on activities related to establishing a new business, farm or professional practice; and/or time spent by a person who owns and manages his/her business or farm even though he/she is physically unable to do the actual work.

Incorporated business (Module P2): refers to the legal status of a business, farm, or in some cases, professional practice. It is directed at persons who were self-employed. An incorporated business is a business or farm, which has been formed into a legal corporation, having a legal entity under federal or provincial laws. An unincorporated business or farm has no separate legal entity, but may be a partnership, family business or owner-operated business.

Industry (Module P2): the general nature of the business carried out by the employer for whom the respondent worked (when an employee, or an unpaid worker in the farm or business of the family), or for their own business (when self-employed).

Job leavers (Module P2): persons who were not working at their job in December 1999 and left that job voluntarily. That is, the employer did not initiate the termination. Detailed reasons

collected are: own health reasons, pregnancy or caring for own children, personal or family responsibilities, going to school, changed residence, dissatisfied with job, found a new job, to concentrate on another job, or another reason.

Job losers (Module P2): persons who were not working at their job in December 1999 and left the job involuntarily, that is the employer initiated the termination. Detailed reasons collected are: company moved, company went out of business, seasonal or non-seasonal layoff, strike, fired, end of contract, or another reason.

Method found job (Module P2): identifies the method used through which the respondent found the job. Methods include: through placement or posting at school, public employment agency, private employment agency, contacted employers directly or sent out resumes, through friends or relatives, placed an add, answered a job add, or through another method.

Net income - ***for self-employed workers*** (Module P2): total earnings for all of 1999 are collected for the self-employed. Net business income is income after all business expenses have been deducted.
(See ***Wages and Salary*** for employee.)

Occupation (Module P2): refers to the kind of work the person was doing at his/her job, as determined by the kind of work reported and the description of the most important duties.

Odd jobs (Module P1): odd jobs may be any type of work for pay and are defined as jobs done on the side to make money, or extra money. These jobs are mostly intermittent such as babysitting, tutoring, yard work, housecleaning, newspaper delivery, etc. **Note**: When a person baby-sits for more than one family, this is considered as one job only. Another example, if a person does many different "odd jobs" to earn extra money, for example, baby-sits and mows neighbours lawns; this is also considered as one job only.

Paid worker (Module P2): someone who works for others (i.e. works for an employer) and receives a wage or salary. The employer usually deducts and remits from the wage or salary income tax, Canada/Québec Pension Plan premiums, etc. There are cases where persons receive a wage or salary but no deductions are made for tax or EI/CPP because the wages earned are too low.

Part-time employment (Module P2): consists of persons who usually work less than 30 hours per week at their job.

Permanent employees (Module P2): permanent employees work at a job for which there is/was no indication that the job would end at some definite point in time. For example, hired permanently with no specified term. (See ***Temporary employee***.)

Reasons for leaving job (Module P2): asked for all jobs that ended prior to December 1999. (See ***Job losers*** and ***Job leavers***.)

Self-employed (Module P2): includes:

- persons for whom the job consisted of operating a business or professional practice, alone or in a partnership. This includes operating a farm whether the land is rented or owned, working on a freelance or contract basis to do a job (e.g., architects, private duty nurses). It also includes operating a direct distributorship selling and delivering products such as cosmetics, newspapers, brushes and soap products, and fishing with own equipment or with equipment in which the person has a share.
- persons who work at "odd jobs" such as occasional babysitting, tutoring, shovelling neighbours entrances.

Temporary employee (Module P2): is an employee for whom there was a definite indication that the job would terminate at some specified point in time. For example, hired for a six month term or a student hired by the same employer during his/her summer holidays or

school breaks. Often referred to as term or contract job by respondents. (See ***Permanent employee***)

Temporary help agency (Module P2): a temporary help agency arranged for the job and the respondent is paid by the agency. For example: Bob does clerical work for Briggs Inc. He obtained this position through Bradshaw Associates, a temporary placement agency. He receives his pay from Bradshaw Associates, not Briggs Inc.

Tips and Commissions (Module P2): paid workers may receive tips, commissions or bonuses in addition to their wage. However, it is likely that the tips, commissions or bonuses are paid on a less frequent basis than the regular wages or salary (e.g. weekly, monthly, etc.). In this case, the value of tips, commissions or bonuses earned are averaged over the period for which the respondent reported their wages or salary. For example, an hourly amount is determined by adding up the total amount of tips, commissions or bonuses received and dividing this by the number of hours worked in that period. This amount is included as part of the hourly rate of pay. (See ***Wages or Salary***)

Unpaid family worker (Module P2): someone who works without pay on a farm or in a business owned and operated by another family member living in the same household. The work done must contribute directly to the operation of a family farm or family business. Respondents who perform regular household chores around the house or yard are not included (e.g. cutting the lawn, painting the house, cleaning the home). (See ***Employee and Self-employed***)

Unpaid leaves from work (Module P2): the term unpaid leave from a job denotes a period of not-working time for which the respondent did not receive any pay from the employer. The period of not-working time was defined, for the purpose of the YITS as four consecutive weeks or longer. The respondent would normally receive a wage or salary from the employer had he/she worked, and may, during the unpaid break receive compensation from some other source such as Workers' Compensation. They are still considered an employee during that time. Unpaid leave periods from jobs were collected from "paid workers", i.e. they were not asked for self-employed jobs and unpaid work in family business or farm jobs.

Wages or Salary (Module P2): for employees, wages or salary is before taxes and deductions (i.e. employment insurance (EI), government pension plans (CPP/QPP), union dues, etc). The respondent chooses the pay period that makes it easier for him/her to give accurate data. For those respondents who choose to report on a yearly basis, the earnings must correspond to an **entire** year, even if the respondent has not worked for the full year (e.g., a respondent started a job a few months ago). The amount entered should reflect what the respondent would normally earn, had he/she worked for a full year.

The category "other" under method of reporting wages and salaries includes persons earning straight commission from their work.

Income from tips, bonuses and commissions are included and averaged over the period for which they apply and included with the reported wages or salary.

"Usual" wage or salary: "Usual" refers to a typical pay period. Where situations are unclear, "usual" pertains to a four week period. If the four week period was not representative of a usual month because the person was on holiday or sick, the respondent is asked for the average earnings under normal circumstances.
(For income of the self-employed, see ***Net Income***)

4.1.7 GAPS - months not in school full-time and not working (Module PS)

For the reference period of 1999, it was possible to determine during the interview the months in which the respondent had not been working or was not in school full-time – referred to as

"gap" months. These "gap" months were derived from the start and end dates of jobs held during the year and from the months during 1999 the respondent was not in school full-time (high school or post-secondary).

For every gap month, the respondent was asked what he/she had been doing, for example, looking for a job to start immediately, going to school part-time, looking after family, etc. For the last gap month, the respondent was asked whether he/she had done anything to look for work, and if so, to name the type of activities.

Collecting information on gap months is very important as it fills in the overall picture of what a respondent had been doing every month within the reference period of 1999.

4.1.8 Volunteering (Module Q)

Volunteer worker: someone who gives his/her unpaid time to a group or an organisation such as charities, schools, religious organisations or community associations. This includes unpaid community service that was done voluntarily, or as a school program, or in order to obtain assistance, or as part of a court sentence. Informal voluntary activities such as painting a neighbour's house or looking after someone's children or pets as a favour is excluded.

4.1.9 Skills (Module R)

Given the changing nature of the workplace and the emphasis on human resource development, the importance of providing skill assessment measures on the YITS was recognised. Due to survey length, such assessment had to be short in duration. Respondents were asked to self-assess (self-evaluate) six skills often used in the workplace, and those generally sought by employers. The skills assessed are ability to use a computer, ability to solve new problems, mathematical abilities, and writing, reading, and oral communication skills.

4.1.10 Income (Module U)

Information collected in the income module is income **received** from all sources **during the year** by the respondent. It is not limited to monies that are taxable.

Employment Insurance : refers to total Employment Insurance benefits received during the year, before tax deductions. It includes benefits for unemployment, sickness, maternity, paternity, adoption, job creation, work sharing, retraining and benefits to self-employed fishermen. As well, include retraining and retirement benefits received under the Human Resources Development Canada (employment insurance program).

Goods and Services Tax Credit (GST) or Harmonized Tax Credit (HST): A person applies for this credit on their income tax return. A person may apply for the credit if they are 19 years of age or older, had a spouse, or were a parent. The credit is based on their net income, added to the net income of their spouse, if applicable.

Income from other government sources (Module U): refers to total income from transfer payments from federal, provincial or municipal governments not reported individually, and received during the year. Included are benefits received under the Canada or Quebec Pension Plan (retirement pensions, survivor's benefits, disability pensions – lump sum death benefits are excluded); Goods and Services Tax credit, and Worker's Compensation.

Income from other non-government sources (Module U): included are sums received from investments, child support, and other non-government sources not reported in separate categories.

> **Income from investments**: includes interest from bank accounts and other deposits, net dividends and other investment income.

Income from child support: includes payments made by a spouse during separation or by an ex-spouse following a divorce; all regular payments and occasional contributions towards the child's maintenance.

Other income: includes income from royalties on books; rental income from other properties; income from roomers and boarders; non-refundable scholarships and bursaries; alimony; and strike pay.

Excludes: tax-free RRSP withdrawals used for purchasing a home, proceeds from the sale of property, businesses, financial assets or personal belongings, loans received and repaid to you as a lender, and refunds of contributions to work-related pension plans.

Money from parents or other people (Module U): included are sums of money (cash) received from a parent or guardian or other people that the respondent does not have to repay. Monies received as loans (regardless of when they are to be repaid) are excluded from income.

Scholarships, grants or bursaries (Module U): income from scholarships, grants or bursaries may be provided either to the respondent or to the institution to pay for tuition. Monies received from any student loans as loans are excluded from income. Money received from fellowships – a position in a university held by a graduate student having teaching duties as part of his or her educational program, is included.

Scholarship: A monetary award to assist a student in the pursuit of studies, based usually on outstanding academic achievement rather than on financial need

Grant: a gift (usually a sum of money) made by a government or corporation (as an educational or charitable foundation) to a beneficiary on the condition that certain terms are accepted or certain engagements fulfilled.

Bursary: a monetary award to assist a student in the pursuit of studies based on financial need and satisfactory achievement.

Self-employment income (Module U): refers to net income (gross income minus expenses of operation such as wages, rents and depreciation) received during the year from the respondent's business or professional practice. In the case of partnerships, only the respondent's share was reported. Also included is net income from persons babysitting, operators of direct distributorships such as those selling and delivering cosmetics, as well as from freelance activities of artists, writers, music teachers, hairdressers, etc.

Social Assistance (welfare) or Provincial Income Supplements (Module U): include payments from provincial or municipal programs for persons in need such as social assistance or welfare for:

- mothers with dependent children
- persons temporarily or permanently unable to work
- the blind and disabled
- benefits covering basic needs (food, fuel, shelter, clothing)
- benefits for special needs (education, respite care, companion services)
- payments from work incentive programs.

Wages and salaries (Module U): refers to gross wages and salaries (including income from commissions, tips and bonuses) before deductions for such items as income tax, pensions and Employment Insurance. Reported are earnings received from all paid worker jobs held during the year.

4.1.11 Personal Characteristics and Family Background (Module V)

Parent's background (Module V): the survey collects information from each respondent about who were the parents with whom they lived with most of the time during high school.

Respondents who reported living with two parents or guardians were then asked if these parents (or guardians) lived together in the same home.

Further questions asked about the parent(s) or guardian(s) highest level of education completed at time of interview and their main job while respondent was in high school.

Information on the number of brothers and sisters was also asked.

Age of respondent (Module V): refers to the age of a person at his or her last birthday as of a specified well-defined reference date (i.e. as of December 1999, at the start of post-secondary education, etc.). Age was derived using the respondent's year of birth that was confirmed during the interview.

Citizenship (Module V): refers to legal citizenship status of the respondent. Persons who are citizens of more than one country were asked to report this information. The concept of citizenship stems from the *Citizenship Act.* Persons may be Canadian by birth and yet hold the citizenship of another country. Persons may also be Canadian by naturalisation and hold citizenship of their country of birth or some other country.

Canada by birth
Persons born in Canada, and those born outside of Canada, if at the time of their birth, one or both parents were Canadian citizens **and** this person has retained Canadian citizenship.

Canada by citizenship process (naturalisation)
Persons who were landed immigrants and have been issued a Canadian Citizenship Certificate are considered Canadian citizens

Other country
Persons who hold citizenship of another country.

Cultural or racial background (Module V): refers to the ethnic or cultural group(s) to which the respondent's ancestors belong.

Dependent children (Module V): children for whom the respondent has sole or joint custody. Included are children for whom the respondent is financially responsible on a regular basis, even if they have infrequent or no contact.

Household (Modules A and V*)*: Household refers to a person or group of persons who occupy the same dwelling and do not have a usual place of residence elsewhere in Canada. It may consist of a family group with or without other unrelated persons, of two or more families sharing a dwelling, of a group of unrelated persons, or of one person living alone. Each person is a member of one and only one household.

Landed immigrant (Module V): persons who have been granted the right to live in Canada permanently by immigration authorities, but have not obtained Canadian citizenship. These persons are referred to as "permanent residents" under the Immigration Act.

Language: capable of working in (Module V): refers to languages in which the respondent can/could use at work. The data are based on the respondent's assessment of his or her ability to speak and /or write.

Language: first learned at home (Module V): refers to the first language learned at home in childhood and still understood by the individual at the time of the survey. This concept relates to ***mother tongue***.

Although respondents may have declared that they learned two or more languages simultaneously, interviewers attempted to have these respondents choose one language over the other. However, in the few circumstances where respondents could not choose between

English and French as their first language, the cases have been included in the derived variable in the category "Other: English and French."

Marital status – youth's (Module V): marital status (conjugal status) of respondent is that of at time of interview. The categories are as follows:

Married
Persons who are legally married, and whose husband or wife are living.

Living with a partner (common-law, girlfriend, boyfriend)
Persons who are not legally married but are sharing the same usual address and living as husband and wife.

Single
Persons who have never married and persons whose marriage has been annulled and who have not remarried.

Separated
Persons currently married, but who are no longer living with their spouse (for any reason other than illness or work), and have not obtained a divorce.

Divorced
Persons who have obtained a legal divorce and who have not remarried.

Widowed
Persons who have lost their spouse through death and who have not remarried.

Province (Entry Module): the data file contains the province that the respondent considers to be their address (variable ndpr); the province in which the respondent last took high school (Module B), and the province where they attended a post-secondary institution (Module H).

Usual Place of Residence (Module V):

School Residence: rooms or apartments that are registered with the institution the respondent is attending.

House, apartment or other private dwellings: single detached dwellings, doubles or duplexes, row or terrace homes, low or high-rise apartments;

Somewhere else: institutions such as penitentiaries, group homes, nursing homes for the aged, hospitals, homes of religious orders, convents; and/or boarding houses, mobile homes, camps, colonies, houseboats, motor homes, hostels, hotels/motels, tourist homes.

Visible Minority (Module V): the concept of visible minority applies to persons who are identified according to the *Employment Equity Act* as being non-Caucasian in race or non-white in colour. Under the *Act*, Aboriginal persons are not considered to be members of visible minority groups.

4.2 Derived Variables

Answers from certain questions from the survey were combined to form specific concepts. Among them: are high school student status; post-secondary school status; and labour force status in December 1999. Such concepts are called: *derived* variables. A derived variable is the result of combining the answers from a number of questions that pertain to a specific concept into a single variable.

The derived variables developed for the YITS have been categorised under education, employment, financing post-secondary, volunteering, demographics, and family background. Within the description of each variable, the variable name is usually in brackets.

4.2.1	Elementary and high school education
4.2.2	Post-secondary education
4.2.3	Financing post-secondary
4.2.4	Employment
4.2.5	Gaps
4.2.6	Activities over each month in 1999 (vector of activities)
4.2.7	Volunteer
4.2.8	Demographics and family background

4.2.1 Elementary and high school education

The derived variables in this section are for respondents who have completed at least some of their education in Canada and the United States, unless specified otherwise.

Date respondent was last in elementary or high school

- Month (HSMD03) and year (HSYD03) respondent was last in elementary or secondary school prior to January 2000.
- Month (LFMD03) and year (LFYD03) respondent was last in elementary or secondary school ***full-time*** prior to January 2000.

For respondents who took some elementary or secondary school in Canada or the United States.

Age of respondent when last in elementary or high school

Two variables give the respondent's age prior to January 2000. Both variables include those who have not yet attained their graduation diplomas.

- Age of respondent who attended elementary or high school ***full-time*** (AFTD03)
- Age of respondent who attended elementary or high school (AHSD03).

For respondents who took some elementary or secondary school in Canada or the United States.

Age of respondent at high school graduation

The derived variable (ACMD03) gives the respondent's age at the date they completed high school diploma requirements or its equivalent.
For respondents who took some elementary or secondary school in Canada or the United States.

Interruptions in elementary and high school education

Variables are for all respondents, whether a high school graduate, continuer or leaver as of December 1999.

- If a respondent has ever dropped out of elementary or secondary school (DRPD03).

For respondents who took some elementary or secondary school in Canada or the United States.
Variable identifies, for all respondents, whether a high school continuer, graduate or leaver as of December 1999, if they have ever dropped out of elementary or secondary school.

- Main reason for having ever dropped out of elementary or secondary school prior to January 2000 (DRED03). For respondents who have dropped out more than once, this variable is the main reason for the last time they have dropped out.

For respondents who took some elementary or secondary school in Canada or the United States. Variable states for all respondents whether they were a high school continuer, graduate or leaver as of December 1999; and the main reason for having ever dropped out of elementary or secondary school. For respondents who have dropped out more than once, this variable is the main reason for the last time they have dropped out.

- Number of times the respondent has ever dropped out of elementary or secondary school prior to January 2000 (DNOD03)

For respondents who took some elementary or secondary school in Canada or the United States. Variable states, for all respondents, whether a high school continuer, graduate or leaver as of December 1999, the number of times they have ever dropped out of elementary or secondary school.

Full or part-time at elementary and high school

Variable (FPTD03) indicates whether the respondent was a full-time or part-time student when ***last*** taking elementary or secondary education.
For respondents who took some elementary or secondary school in Canada or the United States.

High School Status

There are three categories for high school status (HSLD03): high school graduates, high school continuers and high school leavers.

- High school graduates are respondents who have competed the requirements for their high school graduation diploma or its equivalent
- High school graduates are students who have not graduated but whom continue their high school graduations
- High school leavers are respondents who have not graduated from high school and did not complete their high school graduation requirements

For respondents who took some elementary or secondary school in Canada or the United States.

Completion of Grade 9

The variable (HG9D03) shows whether or not the respondent has gone past grade 9 in elementary or secondary school as of December 1999.
For respondents who took some elementary or secondary school in Canada or the United States.

Highest grade in elementary or high school

Highest grade respondent has completed (HGCD03) and, for respondents who have not graduated, the grade enrolled (LGED03) in elementary or secondary school or the date last in school as of December 1999.

- Grade 5 or less
- Grade 6
- Grade 7/ Secondary 1

- Grade 8/ Secondary 2
- Grade 9 / Secondary 3 / Senior 1
- Grade 10 / Secondary 4 / Senior 2 / Level I
- Grade 11 / Secondary 5 / Senior 3 / Level II
- Grade 12 / Senior 4 / Level III
- Grade 13
- SSVD or DEP (where unknown grade equivalence)
- GED/AESS
- Other

For respondents who took some elementary or secondary school in Canada or the United States.

Highest Grade Completed – Main Language and Math Courses

The variables for the highest grade completed for the main language course (HGCL), and the highest grade completed for the mathematics course (HGCM) up to December 31, 1999, are in Module C
For respondents who took some elementary or secondary school in Canada or the United States.

Province

To indicate the province of last high school where the main language and math courses were taken the variables are (PLL and PLM).
For respondents who took some elementary or secondary school in Canada or the United States.

Provincial Level of High School Courses
The variables for the **l**evel of the **l**ast **l**anguage course (LLL) and **l**evel of the **l**ast **m**athematics course (LLM) are also provided for Module C. High School courses are designated under the following headings:

No destination focus for grades listed:
Grades below the high school level and those grades at the high school level where courses are not required as a precondition to a particular destination. Often courses that are required as a precondition to a particular destination begin at higher grades.

University preparatory:
a) Courses which are generally those with the greatest academic challenge, designed primarily for students planning entry into university, into certain programs and colleges and technical schools, or into colleges offering University Transfer programs.

b) Advanced level courses that are enriched and where student achievement in such courses is reported at the advanced level, and assessment is based on comparable standards.

College/Work preparatory:
a) Courses which are less academic and intended primarily for students planning entry into colleges, technical schools, trades or entry into the workplace.

b) Educational experiences which are broadly based and appropriate for all students and which may lead to further studies beyond senior high school years (e.g., apprenticeship, college and university).

Work preparatory:
a) Basic level courses designed for preparation into the world of work. These courses provide a good preparation for direct entry into employment and serve the needs of a student who may not participate in post-secondary education.

For respondents who took some elementary or secondary school in Canada or the United States.

High School Engagement Scales

For module F, variables YSHSCES1were developed to measure overall engagement (YSHSOES1), academic engagement (YSHACES1) and social engagement (YSHSCES1) at high school. The standard error variables for each measure are YSHSCEE1, YSHACEE1, and YSHSOEE1.

For respondents who took some elementary or secondary school in Canada or the United States.

4.2.2 Post-secondary education

Date respondent started post-secondary education

Three levels of start dates have been derived for post-secondary education.

- Month (EDTPSM) and year (EDTPSY) respondent started post-secondary education, prior to January 2000, either on a full-time or a part-time basis.
-
- Month (DSAINM_i) and year (DSAINY_i) respondent started post-secondary education at a particular institution, prior to January 2000.

- Month (DSPRM_ip) and year (DSPRY_ip) respondent started a particular post-secondary program, prior to January 2000.

For respondents who took some education (elementary, secondary school or post-secondary education) in Canada or the United States.

Date respondent ended post-secondary education

Two types of variables have been derived for the post-secondary education end date. One deals only with respondents who have attended full-time programs, while the other includes all respondents taking post-secondary education regardless of their part- or full-time status. This information is provided for three levels of eligible programs.

- Month (DLPSFM) and year (DLPSFY) respondent was last taking post-secondary education on a full-time basis prior to January 2000.
- Month (DLPSM) and year (DLPSY) respondent was last taking post-secondary education ever, prior to January 2000. For post-secondary graduates, these variables give the date of graduation.
- Month (DLFINM_i) and year (DLFINY_i) respondent was last taking post-secondary education at this institution on a full-time basis prior to January 2000.
- Month (DLINM_i) and year (DLINY_i) respondent was last taking post-secondary education ever at this institution, prior to January 2000.
- Month (DLFPRM_ip) and year (DLFPRY_ip) respondent was last taking this post-secondary program on a full-time basis prior to January 2000.
- Month (DLPRM_ip) and year (DLPRY_ip) respondent was last taking this post-secondary program, either on a full-time or part-time basis, prior to January 2000. For post-secondary program graduates, these variables give the date of graduation.

For respondents who took some education (elementary, secondary school or post-secondary education) in Canada or the United States.

Duration of post-secondary program

Time the respondent spent taking a post-secondary program (NMDUR) as of December 1999. Information for this variable is given on a monthly basis.

For respondents who took some education (elementary, secondary school or post-secondary education) in Canada or the United States.

Duration from end of high school to start of post-secondary

The variable (MHSPS) gives the duration of time, in months, from the date the respondent was last in elementary/secondary school to the time the respondent started first post-secondary program. MHSPS is duration in absolute value.

The variable flag (MHSPSFLG) indicates whether duration is negative or positive. A negative value indicates that the respondent started a post-secondary program before elementary/secondary were completed. Likewise, a zero value indicates that the respondent ended elementary/secondary school and started post-secondary education in the same month.

For respondents who took some education (elementary, secondary school or post-secondary education) in Canada or the United States.

Age of respondent starting a post-secondary program

Two variables were designed to capture the respondent's age, in years, when entering a post-secondary program.

- The variable (AGSPS) provides the respondent's age, in years, at the start of their ***first*** eligible post-secondary program.
- The variable (AGEPS) gives the respondent's age at the start of a ***particular*** eligible post-secondary program. For respondents who took some education (elementary, secondary school or post-secondary education) in Canada or the United States.

Interruptions in post-secondary program

Derived variables show if the respondent has ever interrupted their education in a post-secondary program (SIPR_ip), and the main reason for interrupting the program (RSIPR_ip). The variables contain information about programs that are ongoing or completed as of December 1999. Respondents who have stopped their programs before completion (school leavers) are excluded from these variables.

For respondents who took some education (elementary, secondary school or post-secondary education) in Canada or the United States.

Full- or part-time post-secondary student

Variables at the institution (FPLIN_i) and the program (FPLPR_ip) level show if the respondent attended post-secondary education on a full- or part-time basis. Information is collected for the date when last taking post-secondary education prior to January 2000. If two or more programs in the same institution have the same departure date, a respondent who attends at least one of the programs on a full-time basis will be listed in the variable FPLPR as full-time.

For respondents who took some education (elementary, secondary school or post-secondary education) in Canada or the United States.

Post-secondary status by program

There are three categories for post-secondary status by program: post-secondary graduate, post-secondary continuer, and post-secondary leaver (CLGPR_11 – institution 1 program 1 through to CLGPR_41 – institution 4 program 1). In total, up to nine programs were collected for each respondent.

- Post-secondary program graduate is someone who had graduated from their eligible program as of December 1999.
- Post-secondary program continuer is someone who had not graduated from their eligible program as of December 1999, but remained in the program.
- Post-secondary program leaver is someone who has not graduated from their eligible program and who was not in the program as of December 1999.

For respondents who took some education (elementary, secondary school or post-secondary education) in Canada or the United States.

Post-secondary status by institution

There are four categories for post-secondary status by institution: post-secondary graduate-continuer, post-secondary graduate-non-continuer, post-secondary continuer, and post-secondary leaver (HLATT_1 – institution 1 through to HLATT_4 – institution 4).

- Post-secondary institution graduate-continuer is someone who had graduated from one eligible program as of December 1999, and had been in another eligible program at the same institution, in December 1999.
- Post-secondary institution graduate-non-continuer is someone who had graduated from one eligible program as of December 1999, and had not been in any other eligible program at the same institution in December 1999.
- Post-secondary institution continuer is someone who had not graduated from one eligible program as of December 1999, but was in at least one eligible program at the institution in December 1999.
- Post-secondary institution leaver is someone who had not graduated from one eligible program as of December 1999, and was not in any eligible programs at the institution in December 1999.

For respondents who took some education (elementary, secondary school or post-secondary education) in Canada or the United States.

Post-secondary status – overall

Similar to post-secondary status by institution, there are four categories for overall post-secondary status: post-secondary graduate-continuer, post-secondary graduate-non-continuer, post-secondary continuer and post-secondary leaver (LPSAT).

- Overall post-secondary graduate-continuer is someone who had graduated from an eligible program as of December 1999, and had been in another eligible program at the same or in a different institution in December 1999.
- Overall post-secondary graduate-non-continuer is someone who had graduated from an eligible program as of December 1999, and had not been in any other eligible program at the same or in a different institution in December 1999.
- Overall post-secondary continuer is someone who had not yet graduated from an eligible program as of December 1999, but was in at least one eligible program at an institution in December 1999.
- Overall post-secondary leaver is someone who had not graduated from one eligible program as of December 1999, and was not in any eligible programs at the same or in a different institution in December 1999.

For respondents who took some education (elementary, secondary school or post-secondary education) in Canada or the United States.

Highest education level

Highest education level (HEDL) as of December 1999 is divided into four groups: education level below high school completion, high school graduation, some post-secondary education (no certificate), and graduated from a post-secondary program. This variable refers to the highest level of schooling attained by the respondent regardless of whether or not the respondent graduated from this level.

Included in category '1', education level below high school completion, are respondents who attended elementary/high school without completing the requirements for a high school graduation certificate, and those who never attended school.

For respondents who took some education (elementary, secondary school or post-secondary education) in Canada or the United States.

Level of post-secondary program

Level of post-secondary program (LVPR_ip) gives the level for each post-secondary program undertaken by the respondent as of December 1999. The respondent may or may not have graduated from this program.

Category <23> refers to university and non-university level training or programs which are not captured in the other categories.

For respondents who took some education (elementary, secondary school or post-secondary education) in Canada or the United States.

Highest level of post-secondary education

Highest level of post-secondary education (HLPS) taken across all programs and institutions as of December 1999. Respondent may or may not have graduated from this level. The variable's categories are ranked in ascending order <01 to 13>, with the exception of the last two categories '<20> Diploma, certificate or license from a professional association as in accounting, banking, or insurance' and '<23> other level of post-secondary'. In the derived variable, the former category <20> is given preference over the latter <23>.

- Secondary School Vocational Diploma (SSVD)
- Attestation of Vocational Specialisation (AVS or ASP)
- Private business school or training institute diploma or certificate
- Registered Apprenticeship program
- College or CEGEP program
- University transfer program at a college or CEGEP (for credits, university transfer diploma or Associate's Degree)
- College post-diploma or graduate level program (college diploma or higher needed first)
- University diploma or certificate BELOW Bachelor's (undergraduate level)
- Bachelor's degree
- First professional degree
- Graduate-level diploma or certificate ABOVE Bachelor's, BELOW Master's
- Master's degree
- Ph.D. degree
- Diploma, certificate or license from a professional association as in accounting, banking, or insurance
- Other level of post-secondary

For respondents who took some education (elementary, secondary school or post-secondary education) in Canada or the United States.

Highest graduation diploma completed

The highest graduation diploma attained by the respondent, as of December 1999, consists of respondents who did not graduate from high school or a post-secondary program, those who graduated from high school, and those who graduated from a post-secondary program. Two variables were created to capture the highest diploma completed based on where the respondent attended school.

- The first derives information solely for respondents who have taken at least some of their education in Canada and the United States (HGDA).

- The second is for all respondents, including those who attended school in Canada, the United States or elsewhere (HGDAA).

Highest education level completed

Highest certificate, diploma or degree ever attained (HEDAT) as of December 1999. Respondents who have attended high school but did not graduate are listed as a valid skip.

- High school graduation requirements completed
- Post-secondary non-university diploma or certificate (*i.e. a college or CEGEP certificate or diploma, registered apprenticeship programs, private business school or training institutes, and Quebec's Secondary School Vocational Diploma & Attestation of Vocational Specialisation)*
- Undergraduate certificate/diploma below Bachelor's degree, or University Transfer Program
- Bachelor's Degree
- First Professional Degree (e.g. medicine, dentistry, veterinary medicine, law, optometry, divinity)
- University Degree/Diploma above Bachelor's level
- Master's Degree
- Doctorate (PhD)
- Other Programs *(includes a diploma, certificate or license from a professional association as in accounting, banking or insurance, as well as other programs above the high school level not included in the other categories).*

For respondents who took some education (elementary, secondary school or post-secondary education) in Canada or the United States.

Major field of study

Major field of study refers to the predominant discipline or area of learning or training of a person's post-secondary degree, certificate or diploma. The major field of study classification structure consists of 10 broad or major categories: educational, recreational and counselling services; fine and applied arts; humanities and related fields; social sciences and related fields; commerce, management and business administration; agricultural and biological sciences/technologies; engineering and applied sciences, engineering and applies science technologies and trades; health professions; sciences and technologies; and mathematics and physical sciences. This structure is, in turn, subdivided into over 100 "minor" classifications categories and about 850 "unit" groups.

Major field of study (1P1MFS1, I1P2MFS1, I1P3MFS1, I2P1MFS1, I2P2MFS1, I2P3MFS1, I3P1MFS1, I3P2MFS1, I4P1MFS1) have been derived for the major field of study reported for each eligible program reported by the respondent. Respondents were also able to report a second major field of study, if the respondent felt it was of equal importance (1P1MFS2,

I1P2MFS2, I1P3MFS2, I2P1MFS2, I2P2MFS2, I2P3MFS2, I3P1MFS2, I3P2MFS2, and I4P1MFS2).

For respondents who took some education (elementary, secondary school or post-secondary education) in Canada or the United States.

Order of post-secondary programs

The order of post-secondary programs attended by the respondent is given in two variables.

- The first post-secondary institution and program ever attended (FPSP).
- Order of all post-secondary institutions and programs attended by respondent (OPSP). The derived variable orders the post-secondary programs a respondent has taken from the earliest (first one taken) to the most recent.

For respondents who took some education (elementary, secondary school or post-secondary education) in Canada or the United States.

Number of eligible programs and institutions

All programs were checked for eligibility. To be eligible, a program must be above the high school level, towards a diploma, certificate, or a degree, take at least three months to complete, and started prior to January 2000.

- Number of eligible post-secondary programs taken by the respondent as of December 1999 (NPRDI).
- Number of post-secondary institutions attended of December 1999 (NINDI).
- Number of eligible post-secondary programs per institution (NPRPI_1:4). The variable refers to the number of programs taken by the respondent in each institution, where the number of programs does not exceed the limit. Thus, 3 programs are collected for institutions 1 and 2, 2 programs for institution 3 and 1 program for institution 4.

To be eligible, a program must be above the high school level, towards a diploma, certificate or degree, take at least three months to complete, and started prior to January 2000.

For respondents who took some education (elementary, secondary school or post-secondary education) in Canada or the United States.

4.2.3 Financing post-secondary education

Total amount of money owed to student loans, bank loans, lines of credit, parents or family as of December 1999, to fund post-secondary education (TMOD010).

Total amount of money received from scholarships, awards, prizes, grants or bursaries as of December 1999, to fund post-secondary education (TMRD010).

For respondents who took some education (elementary, secondary school or post-secondary education) in Canada or the United States.

4.2.4 Employment

Average number of hours worked per week by students

The average number of hours worked per week while taking some education was asked of first-year post-secondary students and students taking their final year of high school. The range of values for both sets of variables is similar.

- Average number of hours of paid work each week during the last year of high school (HPDHS).
- Average number of hours of unpaid work in family's business or farm during the last year of high school (HUWHS).
- Average number of hours of paid work each week during the first year of post-secondary education (HPDPS).
- Average number of hours of unpaid work in family's business or farm during the first year of post-secondary education (HUWPS).

For respondents who took some education (elementary, secondary school or post-secondary education) in Canada or the United States.

First job after full-time schooling

YITS collects information on jobs worked in 1999 (maximum of six jobs), as well as on the "First job after leaving full-time schooling" (also referred to as the 7th job). For this job, information is collected even if the respondent did not work at it in 1999, but prior to that date.

The 7th job (first job after leaving full-time schooling) was only collected in the case where the respondent had reported more than 4 jobs and wasn't at school full-time in December 1999. This job was also collected automatically in the case where the respondent had stopped full-time schooling prior to 1999.

- The derived variable (FSTJOB) identifies the first job after leaving full-time schooling among the different jobs reported by the respondent in Cycle 1 Main Survey. This job could have either been already specified by the respondent during the interview or will have to be chosen from a set of valid jobs reported by the respondent.
- The duration, in months, between time respondent left full-time schooling and time respondent started first job (FTSFJD). A negative value for this derived variable indicates that the respondent started their first job prior to leaving full-time schooling. A positive value indicates that the respondent started their first job after leaving full-time schooling (FTSFJFLG).
- Age of respondent when they started first job after full-time schooling (FJ_AGE). In a few circumstances, the age of the respondent may vary by one year, due to the nature of the annual information on which the derived variables are based.

For respondents who took some education (elementary, secondary school or post-secondary education) in Canada or the United States.

Number of jobs in 1999

Eligibility (see definitions in section 4.1.6) plays a significant role in the derivation of number of jobs the respondent worked at during 1999. First, the verification of job eligibility is not performed on all jobs, but on a maximum of the first six jobs reported by the respondent for the reference period 1999. In other words, a respondent with *n* jobs in 1999 where *n* is greater than 6, has (*n* – 6) jobs whose eligibility could not be verified. Second, if a respondent's first job after leaving full-time schooling was different from the jobs reported for 1999, then this particular job, also known as the "seventh" job, is automatically considered

eligible. So a respondent can have a maximum of 6 jobs (in 1999) + the "seventh" job (first job after leaving full-time schooling) for a total of 7 jobs whose eligibility has been verified.

- The derived variable that represents the number of jobs in 1999 where the eligibility has been verified is (ELGJOBV).
- The derived variable that represents the number of jobs where the eligibility could not be verified [(n – 6) jobs in 1999 where n > 6] is (ELGJOBNV). Because the eligibility of these jobs could not be verified, users should be cautious when using this variable. The total number of jobs in 1999 is equal to the sum of (ELGJOBV) and (ELGJOBNV).
- The order of jobs respondent worked at (OJOB) is given from the earliest job begun to the job started the most recently.

Work Hours

The derived variables are for each job held by the respondent.

- Number of paid hours usually worked per month (NHWPMS_j) when respondent first started working at the job (for each listed job).
- Number of paid hours usually worked per month (NHWPME_j) when respondent was last at the job (for each listed job).
- Number of weeks usually worked per month (NWWPMS_j) when respondent first started working at the job (for each listed job).
- Number of weeks usually worked per month (NWWPME_j) when respondent last worked at the job (for each listed job) prior to January 2000.
- Indicator for whether the respondent usually worked 30 or more hours per work when last at job (HWE_j).
- Indicator for whether the respondent usually worked 30 or more hours per work when first started at the job (HWS_j).

Wages and Salaries

All variables depicting earnings include tips and commissions.

- Earnings when first started job (for each listed job) – Hourly (EPHS_j)
- Earnings when first started job (for each listed job) – Weekly (EPWS_j)
- Earnings when first started job (for each listed job) – Monthly (EPMS_j)

- End earnings when last worked at job (for each listed job) – Hourly (EPHE_j)
- End earnings when last worked at job (for each listed job) – Weekly (EPWE_j)
- End earnings when last worked at job (for each listed job) – Monthly (EPME_j)
- Total earnings in 1999 from all self-employed work (for all listed jobs) (EARNES_j)

Job Tenure

Three variables have been derived about the duration, in months, of each job held by the respondent.

- Total number of months between time the respondent started working at job j and the time respondent last worked at job j as of December 1999 (TENURE_j). The respondent is considered as being employed at job j during unpaid leaves from the job.
- Total number of months the respondent was employed at job j **during 1999** (TNUR99_j). Respondent is considered as being employed at job j during unpaid leaves from the job.
- Total number of months the respondent was employed and worked at job j during 1999 - less paid and unpaid leaves (NMW99_j). In this variable, respondent is **not** considered as having worked at job j during unpaid leaves from the job.

P85 Main activity during the period of one month or more in between the time respondent left full-time schooling and the time s/he started his/her first job after leaving full-time schooling.

Labour market activity: Occupation
Standard Occupational Classification (2 digit)

For purposes of identifying variables for the codebook – all Occupational Codes (4 digits) were renamed from "SOC" to "OCC.

Respondents were asked to describe the kind of work they were doing and their most important activities or duties. The information from these responses was then used by coders to assign an occupation code from the 1991 Standard Occupational Classification (SOC).

The most detailed SOC code has 1 character, followed by 3 digits[3].

This classification is composed of four levels of aggregations. There are 10 broad occupational categories containing 47 major groups that are further subdivided into 139 minor groups. The minor groups contain 514 unit groups.

If the respondent did not specify an occupation or did not define it in sufficient detail to permit coding then a code of 11 was assigned.

For the YITS, the SOC code has been derived at the 2 digit broad occupational level for:
- the respondent's occupation for up to a maximum of 4 jobs (JOBROC_j – 4 digit occupation code; JOBOCC_j – 2 digit occupation code)
- the parents/guardians' (with whom the respondent lived most of the time during high school) main job while the respondent was in high school – to a maximum of four (PAROC – 2 digit occupational code; PAOCC – 4 digit occupational code). These variables are not gender specific
- the job the respondent would like to have when he/she is 30 years of age – 4 digit occupation code (FJOBROC)
- the job the respondent would like to have when he/she is 30 years of age – 2 digit occupation code (FJOBOCC). See "Ever Worked at Job or Business" below.

Labour market activity: Industry
North American Industry Classification System (2 digit)

All Industry Codes (4 digits) were renamed from "NAICS" to "IND".

Respondents were asked to indicate for whom they worked (their employer name) and the kind of business or service offered by this employer. The information from these responses was then used by coders to assign an industry code from the 1997 North American Industry Classification System (NAICS).

If the respondent did not specify an industry or did not define it in sufficient detail to permit coding then a code of 17 was assigned.

The most detailed NAICS code has 4 digits.[4] For the YITS, the NAICS code has been derived at the 2 digit industry sector level for:
- the respondent's industry for up to a maximum of 4 jobs (JOBRIC_j – 4 digit industry code; JOBIND_j – 2 digit industry code)
- the parents/guardians' (with whom the respondent lived most of the time during high school) main job while the respondent was in high school for a maximum of 4 jobs (PARIC – 2 digit industry code; PAIND – 4 digit industry code). These variables are not gender specific.

(See "Ever Worked at Job or Business" below.)

[3] In the SOC code, the first character identifies the broad occupation category, the second digit identifies the broad occupational group, the third digit identifies the minor occupational group, and the fourth digit identifies the unit group.
[4] In the NAICS code, the first two digits identify industry sector, the third digit identifies industry sub-sector, and the fourth digit identifies industry group.

4.2.5 GAPS – Months not in school full-time and not working

Last Gap Month

A gap month is a month during the reference period where the respondent was not employed at a job and was not a student full-time. The 'last gap month' is the last month during the reference period where this occurred. Respondents were asked if they did anything to look for work, and what activities they did to look for work, during their last gap month only. The reference period for YITS cycle 1 is January 1999 to December 1999.

The variable (LGM) indicates which month in the reference period was the respondent's last month. The variable (LWLGM) indicates, for the respondent's last gap month, if he or she was looking for work.

Monthly post-secondary status

Derived variables have been created for each of the twelve months in 1999 to indicate post-secondary student status.

- A variable (FEDST_mm) indicates for which month(s) the respondent attended a post-secondary program full-time. Coverage for all respondents

- A variable (FTES) indicates whether the respondent attended full-time elementary/secondary school or a post-secondary program for each month of 1999. For this variable, both part-time and full-time post-secondary students are included. For respondents who took some education (elementary, secondary school or post-secondary education) in Canada or the United-States.

Ever worked at a job or business:

NOTE: For purposes of identifying variables for the codebook – all Occupational Codes (4 digits) were renamed from "SOC" to "OCC". All Industry Codes (4 digits) were renamed from "SIC" to "IND"

Derived variables were based on questions asked of respondents who did not have a job in 1999 and did not have a first job after leaving full-time schooling.

Respondents who were:

1. full-time students in December 1999 and did not work at a job during 1999; not full-time students in December 1999, did not work at a job during 1999, and indicated not having had a job after leaving full-time schooling; or
2. whose work history for 1999, or whose first job after leaving full-time schooling is unknown;

AND respondents who were asked if they have ever worked at a job for pay, at odd jobs for pay, or work in their family's farm or their family's business with or without pay.

From these questions, it was possible to derive variables for:

- Respondents who have ever worked at a job or a business (EVWRK) in 1999.
- The class of worker (EMPST): whether an employee, a self-employed worker, an unpaid worker in family business or farm
- The type of worker (PTFTST): either full-time (full-time employment consists of persons who usually work 30 hours or more per week at their job) or part-time (part-time employment consists of persons who usually work less than 30 hours per week at their jobs).

Respondents were asked to indicate for whom they worked (their employer name) and the kind of business or service offered by this employer. The information from these responses was then used by coders to assign an industry code from the 1997 North American Industry Classification System (NAICS). A breakdown by 2-digit (EWRIC) and 4-digit (EWIND) NAICS codes are given. If the respondent did not specify an industry or did not define it in sufficient detail to permit coding then it was assigned a code for 'unclassified' (17 or XXXX respectively).

Respondents were also asked to describe the kind of work they were doing and their most important activities or duties. The information from these responses was then used by coders to assign an occupation code from the 1991 Standard Occupational Classification (SOC). The most detailed SOC code has 1 character, followed by 3 digits. This classification is composed of four levels of aggregations. There are 10 broad occupational categories containing 47 major groups that are further subdivided into 139 minor groups. The minor groups contain 514 unit groups. A breakdown by 2-digit (EWROC) and 4-digit (EWOCC) SOC codes are given. If the respondent did not specify an occupation or did not define it in sufficient detail to permit coding then a code of 17 or XXXX was assigned.

Reservation wage

Reservation wage is the lowest wage or salary a respondent would accept to begin a new job immediately, full-time. It is also known as a worker's asking wage. This variable is expressed in dollars and cents per hour. (RSWG)

Outlier detection and solution:
Using standard methodology, 8 observations (0.04% of the total sample) were observed to have extreme or nonsensical reservation wage. A solution was to set these values to $1000.00, the maximum value of the reservation wage just before the observed extreme values.

Employment Status

The Employment Status can be derived from a maximum of 7 eligible jobs as the start and end dates are collected for a maximum of 7 jobs in the P1 Module. From this total, details are collected on a maximum of 5 jobs (first 4 jobs + first jobs after leaving full-time schooling (seventh job)) in the P2 Module. Situations where the respondent is not at school full-time and is not working are collected from the GAPS Module. The status has been derived for each eligible job and for the respondent overall.
Start dates (JSTDAT_7) and End Dates (JENDAT_7) of jobs worked in 1999 are written for the first job worked at after leaving full-time schooling.

To indicate whether job 7 corresponds to either job 5 or 6 (JOB7FLG)

Job status as of December 1999

Job status refers to whether or not the respondent had a job in December 1999. The status has been derived for the month in which the respondent had a job (JBSTmm)

Working status as of December 1999

It is possible that a respondent had a job in December 1999, but did not work at that job (i.e., because of an unpaid leave). Therefore, the working status as of December 1999, whether the respondent was working at the job(s) he/she held in December 1999, has been derived for each eligible job (WKSTmm_j(1:7), A respondent may have been employed at more than one job and the information on unpaid leaves was ignored if the data was incomplete (WKSTmm)

Users should note that:

- Job details were collected only for the first four jobs in 1999, as well as for the first job after leaving full-time schooling. This means that the working statuses for jobs 5 and 6 are unknown.
- Also, if a respondent did not have a job in December 1999, then working status is non-applicable.

Working full-time or part-time as of December 1999

For those respondents who were working at a job(s) in December 1999, their full-time / part-time work status has been derived for each eligible job (WKFTPT_(1:4, 7), and for the respondent overall (WKFTPT).

Users should note that:

- Job details were collected only for the first four jobs in 1999, as well as for the first job after leaving full-time schooling. This means that the full-time/part-time work statuses for jobs 5 and 6 are unknown.
- Also, if a respondent did not have a job in December 1999, or the respondent did not work at the job in December 1999, then full-time/part-time status is non-applicable.

Main activity of respondent who did not have a job in December 1999

For those respondents who did not have an eligible job in December 1999, their main activity has been derived for each month when not employed at a job (REASONmm). For every month in 1999 where a respondent was not employed at a job and was not a full-time student, s/he was asked what their main activity was during the month.

The variable WKFTPT_j (1:4-7) was used to create the full-time and part-time employment status for each job. As this variable was derived using WKST12_j, the information on unpaid leaves was ignored if the data was incomplete, either because the respondent did not indicate whether any unpaid breaks occurred, how many breaks occurred, or did not indicate the starting or ending month of a break.

Job Change

The job change variables show whether the respondent has ever left one eligible job and started another. One variable determines for each month in 1999 whether the respondent changed jobs (CJ-mm), while the other indicates if the respondent changed jobs during one month at any time in 1999 (CJ_1999).

Multiple Jobs

Multiple jobholders refer to respondents who were employed at two or more jobs during 1999. The status has been derived for each month (January to December) in 1999 (MLTJB_mm). An overall summary of MLTJB_mm is given to indicate if a respondent was employed at multiple jobs during one month at any time in 1999 (MLTJB_Y).

4.2.6 Activities over each month in 1999

The vector of work-status activities is designed for every month in 1999 when the respondent stated <1> Had a job during the month; <2> No, did not work; or <6> did not work at all in 1999.

Employment status flag (JBSTmm) for each month in 1999 indicating if respondent was employed at at least one job - for jobs 1 to 7. The variable indicates if the respondent was employed at at least one job during the month. Note, during a given month, respondent may have worked or may not have worked at a job (in situations where respondent is on an unpaid leave from the job).

A 2 digit work date variable (WRKDTE_mm) is created for all months from 01 (January) through to 12 (December).

4.2.7 Volunteer

Overall volunteer activity

This variable identifies whether or not a respondent had done any volunteer activities during 1999 (OVRD020).

4.2.8 Income

Income Sources

Income variables from Module U have been produced for each of the income sources (INCEI – employment insurance; INCJB – wages and salaries; INCNG – non government sources; INCNR – parents or others; INCOG – other government sources; INCSA – social assistance or welfare benefits; INCSGB – scholarships, grants, bursaries)

Total Income

Total income is a sum from all sources of income indicated by the respondent (TINCD – all sources)

4.2.9 Demographics and family background

Age in December 1999

Using the respondent's year of birth that was confirmed during the interview, the respondent's age as of December 1999 was derived (AGEDV). This variable is not released in the codebook.

Location of resident's household
Several variables depict the geographical unit of the respondent's residence for the household as of date of interview (1996 Census Geography).

The only two geographical variables available on the data file will be:
- PROVDV: Province of residence for the household as of date of interview
- URBRURMZ: Indicator of rural or urban geography, based on the MIZ coding (Metropolitan Influence Zone), based on the 1996 Census geography

For confidentiality purposes, detailed geography variables are not available in the YITS data files. The geographical variables noted below may be made available upon request by researchers or analysts. Statistics Canada will evaluate each request on a case by case basis in relation to what the researcher is planning to release and the files to which they are planning to link.

Each variable represents the address confirmed for the respondent from the question Z1. Personal address information given by the respondent may not reflect where the respondent lives throughout most of the year. For example, students who move away from home during the school year to attend a post-secondary institution may provide either own address while at school, or their parent's residence.

Respondents with non-Canadian addresses are included in category <996> as a valid skip.

- REG**:** Region of residence for the household as of date of interview
- CMA25: Census Metropolitan Area (CMA) of residence for the household as of date of interview
- CMACA: Census Metropolitan Area (CMA) or Census Agglomeration (CA) of residence for the household as of date of interview

- CSDC: Census Subdivision Code (CSD) of residence for the household as of date of interview (1996 Census Geography)
- ECR: Economic region (ER) of residence for the household as of date of interview
- SATYPE: Statistical Area Classification (SAC) type of residence for the household as of date of interview. The categories are: CMA, CA, CA (no Census tract), strong MIZ (within provinces), moderate MIZ (within provinces), weak MIZ (within provinces), no MIZ (within provinces) and CSD within a territory.
- EIER: Employment Insurance Economic Regions (1996 Census Geography)
- EAUID: An enumeration area (EA) is the geographic area canvassed by one census representative. It is the smallest standard geographic area for which census data are reported. All the territory of Canada is covered by EAs (1996 Census Geography)

Working languages

Based on the respondent's own assessment of his/her ability to speak, we have derived the languages which the respondent feels capable of working in (LCWD23). The groupings are divided into English, French, other non-official languages or combinations of two or more languages.

Referring to the first language learned at home in childhood and still understood by the individual at the time of the survey, also referred to as mother tongue (LFLD23)

Respondents were asked about the first language that they first learned at home in childhood and still understand. This concept relates to mother tongue (V7)

Although respondents may have declared that they learned two or more languages simultaneously, interviewers attempted to have these respondents choose one language over the other. However, in the few circumstances where respondents would not choose between English and French as their first language, the cases have been included in a new category under variable V7 "Other English and French".

Visible minority

The concept of visible minority applies to persons who are identified according to the *Employment Equity Act* as being non-Caucasian in race or non-white in colour. Under the *Act*, Aboriginal persons are not considered to be members of visible minority groups.

The variable (VMSD23) indicates whether the respondent is a visible minority defined by the *Act*.

Number of years in Canada

The number of years since the respondent's first coming to Canada to live, as of December 1999 (YFCD23).

Age at relocation to Canada

Respondent's age at time of moving to Canada (ATMD23). In a few cases, the age of the respondent may vary by one year, due to the nature of the annual information on which the derived variables are based.

Age leaving family home

Respondent's age, in years, at time of moving out permanently from the home of their parents or guardians (AGMP). In a few cases, the age of the respondent may vary by one year, due to the nature of the annual information on which the derived variables are based.

Parent/Guardian structure

In the family background section of the YITS, respondents were asked to list all the parents and guardians they lived with most of the time during high school. This information was used to derive the parent/guardian structure variable (FAMSTAT).

The following table was designed to define the relationship in the household of the YITS 18-20 year-olds and the PISA/YITS 15 year old reading cohort (parent file):

PISA Family Structure		YITS Cohort B (18-20 year-olds)
FAMSTRUC	**Condition**	**FAMSTAT**
1 Single	if only one of STQ4a, STQ4b, STQ4c or STQ4d is YES.	21, 22, 23, 24 25, 26, 27 or 28
2 Nuclear	if both STQ4a and STQ4c are both YES, and if both STQ4b and STQ4d are both NO	01 or 04
3 Mixed	if STQ4a and STQ4d are both YES or if STQ4b and STQ4c are both YES.	02, 03, 05, 06, 07, 08, 09, 10
Missing	If all STQ4a, STQ4b, STQ4c and STQ4d are missing	13, 29, 34, 41, 42, 97, 98, 99
4 Other	Else	11, 12, 31, 32, 33, 42, 50

Note 1:
For the derived variable with values of 13, 25, 26, 27, 28, 29 and 34, “other” includes grandmother, grandfather or other non birth, step, adoptive, foster parent/guardian.
Note 2:
The derived variable with a value of 50, may include biological parent(s) and/or guardian(s) possibly living together in the same home or living separately.

Parents/Guardians in the Family Home (PGatH)

- respondent lived with selected parents during high school
- respondent had parents in family home while away during high school

Parent/Guardian at Home

The derivation of this variable required that certain decisions be made. Users should note the following:

In defining family structure, biological parents took precedence over other parent or guardian figures. For example, if a respondent indicated that they lived with both biological parents and that both biological parents lived in the same house, they were coded as value “01". This was despite the fact that some respondents may have also reported another parent or guardian figure such as grandmother living in the same household. (PG_1:4)

In split custody situations, it was possible that more than two parent or guardian figures were identified. In these situations, it was not possible to determine who lived in which household. For example, respondents may have reported two birth parents and two stepparents. Family structure, in these cases, was determined by the presence of the biological parent. So, for example, in the case of two birth parents and two step parents, family structure would take on the value of (31) split custody - two biological parents. The split custody variable can only be used to assess the presence of two biological parents. It is possible that two biological parents lived in the same household and that the other person identified (a foster mother for example)

lived in the other household. It cannot always be presumed that the biological parents are in two separate households.

"Other guardian" refers to a person who was not a biological, adopted, step or foster parent. This includes other relatives such as sisters, brothers, aunts and uncles as well as non-family members such as the boyfriend/girlfriend of the parents, etc.

One parent families who have another guardian figure such as a grandmother, grandfather or an "Other specify" are also identified in values 25 through 28. From an analytical point of view, this may be helpful in determining if having support of another adult in the household has an impact on one-parent families. However, it should be noted that the question was not designed to solicit specific information on one-parent families with support. It is possible that a respondent simply listed all adult figures in the household and that these cases represent part of the "sandwich" generation where parents are looking after both their children and their own parents at the same time.

Parental/Guardian Education

The level of education for each person listed as a parent/guardian (up to a maximum of 4 persons) (PED1:4)

4.3 YITS Scales

Some of the concepts in YITS cycle 1 are most appropriately measured through the use of a scale. For the YITS 18-20 year old cohort, only one scale was used (the high school engagement scale). For more theoretical details about any of the mathematical / statistical concepts discussed in this section, please see the Statistics Canada technical document *Analysis of Scales for YITS/PISA Cycle 1 Main Survey* which should be available in the winter of 2005.

This section is divided up into five major subsections. The first subsection **4.3.1 Defining Scales and Their Uses** introduces the user to the concept of a scale, addresses the justification for scaling and describes the type of scaling applied within YITS. The second subsection **4.3.2 Scale Development** discusses the theoretical procedure that the YITS team went through to form a scale. The third subsection **4.3.3 The High School Engagement Scale** introduces the user to the concept of school engagement, the High School Engagement Scale, and to its subscales. This subsection also provides justification for why the YITS team felt that it was important to create a High School Engagement Scale. The fourth subsection **4.3.4 Analysis of the High School Engagement Scale** discusses the analysis involved in the formation and testing of the High School Engagement scale. This involved applying the methodology introduced in subsection **4.3.2** to the theoretical model discussed in **4.3.3**. The last subsection **4.3.5 References** contains a list of references cited or consulted within the previous four subsections.

4.3.1 Defining scales and their uses

What is a Scale?

A scale is a type of composite measure consisting of several items that share an empirical or logical structure. A scale can be regarded as a set of possible scores for combining answers to a group of questions (items).

Why Use a Scale?

The use of scales in data analysis allows researchers to measure a particular underlying concept when these items are put together. A scale is created by assigning scores to patterns of responses that enable the analyst to assess the relative weakness or strength of each indicator. The use of scales is advantageous in that scales can demonstrate the intensity of the relationship that may exist among attributes.

For each factor (concept) measured by a scale, a latent score value is estimated for each individual surveyed within the sample of eligible respondents. This estimated score is based upon appropriate combinations of a number of items (responses to a group of survey questions). The score for a particular factor may be used to order individuals with reference to the factor or to illustrate differences between individuals or groups with respect to that factor.

A scale has a higher level of reliability than do individual items. An item can contain confounding variance (noise) due to measurement error and information uniquely associated with that item as well as information about the construct being measured. Using a scale helps to reduce this noise. Scales are useful in social science research because they facilitate the efficient reduction of large amounts of data into manageable and meaningful sources of information for the analyst (Barbie, 1995, cited in *NLSCY scales*).

What Type of Scales are Used in YITS?

The various scales used in YITS (both in the PISA/YITS Reading cohort and in the YITS 18-20 year old cohort) are modelled after the Likert Scale (Likert, 1932). This type of scale is valued for the ordinality of its multiple response categories. This allows researchers to judge the relative strength of agreement that was intended by the respondents of the survey to any particular item. For example, a particular question with four categories may require respondents to choose from four ordinal scale values such as 1 (strongly disagree), 2 (disagree), 3 (agree) and 4 (strongly agree) that describe the respondent's view on the item prompt.

A Likert-type scale is a highly reliable tool for rank-ordering people when attempting to measure certain attitudes or opinions that they may hold. It is one of the most commonly used and most easily recognizable formats in the area of questionnaire design. Likert scales can have any number of categories. Indeed, the more categories provided to a respondent, the more precise the distinction between the categories and the more information available to distinguish between respondents on the measured construct. However, in practice, respondents may not be able to respond meaningfully when there are too many categories. Moreover, there is additional burden to the respondent in terms of the time required to make such fine distinctions. The application of scale-type questions is a very inexact science; their use is somewhat subjective. The number of item categories should be chosen based upon situation-specific judgment including knowledge of the item content, the underlying scale construct, and the respondents.

In many of the items that make up the scales, a neutral category such as *neither agree nor disagree* or *sometimes* has been excluded. Many of the questions asked in YITS scales deal with topics that are not often considered by respondents; thus, respondents could have tended toward a neutral response, which would have reduced the utility of the scale. Removing the neutral category, in this case, makes it more likely that the scale will detect tendencies of respondents, even if they are slight. Some experts in questionnaire design feel that if a respondent does not know whether they are leaning to the positive or the negative end of a statement that they should indicate "don't know" rather than indicating a neutral response. Other experts, however, think that the neutral response is necessary for those respondents who truly do not have an opinion on the statement. The theory and design of survey scale items is discussed in more detail in *Survey Research Methods* (Fowler, 1995).

In order to meet the specific survey objectives of YITS (**section 3.0)**, a scale was constructed by the YITS team to measure high school engagement. Standard models of school engagement from previous research were modified through expert consultation for administration in YITS. Validation of chosen items was undertaken using data from the pilot and main survey of YITS conducted in 1999 and 2000 respectively. A general description of how YITS scales were developed for both the PISA/YITS Reading cohort and the YITS 18-20 year old cohort data (**section 4.3.2**) as well as the particularities of the development and validation of a high school engagement scale for the YITS 18-20 year old cohort are detailed below (**section 4.3.4**).

4.3.2 Scale development

There were three main steps in the development of the scales. The first step was to use linear factor analysis to investigate if the underlying theoretical structure of items and sub-scales was supported by the data (**section 4.3.2.1**). Assuming that the theoretical structure was valid, the second step in development was to estimate the scale scores using an item response theory (IRT) model (**section 4.3.2.2**). Finally, reliability and validity checks were performed on the estimated scores (**section 4.3.2.3** and **section 4.3.2.4**). The general procedures that were adopted for each of these steps are described below.

4.3.2.1 Investigation of model validity

Factor Analysis: Strategy

Factor analysis was used to determine whether the theoretical construct of the scale was supported by the data. Moreover, factor analysis is one way to help achieve the goals of an item analysis that is the verification that items are related to their proposed constructs and that the strength of these relationships were adequate for measurement purposes. (Gorsuch, 1997)

There is some debate within the factor interpretation community about the meaning that can be applied to a factor. This controversy centers around whether the latent structure determined by a factor analysis truly exists as a "real" and meaningful structural aspect or whether it is merely a mathematical edifice (an arbitrary descriptive variable). Comrey and Lee provide the view that the goal of factor analysis is to isolate constructs that have a greater intrinsic value for the purpose of describing the functional relationships between the variables in the field. However, all sets of variables are not equally good in representing this relationship and there is not necessarily only one "correct" concept or "real" factor for a given domain of interest. (Comrey & Lee, p. 245, 1992)

The relationship of each variable to each of the factors, referred to as the loading on a factor, provides a way to quantitatively assess how an item interacts with other items. The strength of these loadings on different common factors indicate whether an item is related to none of the proposed factors, to only one of the factors or to more than one factor. The greater the loading on a factor, the greater the amount of overlapping (common) variance[5] between the data variable and the factor and the more the item is a pure measure of this factor.

Items most clearly associated with only one factor can then be recommended as a scale for the construct underlying that factor. Items not strongly associated with any of the factors can be dropped from further analysis. Sometimes an item may be strongly associated with more than one factor. These items may indicate a relationship between the two factors, in which case many items should load on more than one factor and the double loading can be explained mathematically. Alternately, they may indicate the presence of an unknown or confounding factor that is related to both factors, in which case the item or items in question should be dropped from the scale.

A question that often arises when examining factor loadings is how high the correlation between an item and a factor must be before the item is considered as "significant". Although no formal test to determine significance has been developed, Comrey and Lee provided a crude benchmark to use in interpreting variance-factor correlations[6]. They

5 The variance in responses to any particular item can be described by two main components: the specific variance and the common variance. Specific variance represents the differences between people related to the unique characteristics of the item. Common variance, or communality, refers to differences that can be explained by the common factor related to all items in the scale.

6 Orthogonal factor loadings or structure coefficients

base their benchmark upon an examination of the percentage of variance of the item common to the factor (See **Table 4.3.1**).

Table 4.3.1 Scale of Variance-Factor Correlations

Loading	Percentage of Variance Explained	Quality of Loading
Above 0.71	Above 50	Excellent
Above 0.63	Above 40	Very Good
Above 0.55	Above 30	Good
Above 0.45	Above 20	Fair
Above 0.32	Above 10	Poor
Below 0.32	Inconsequential	Trivial (not to be interpreted)

Examining factor loadings, however, is not in and of itself sufficient for factor interpretation. A full explanation of a factor requires a theoretical understanding of why the items fit together and how the identified factor or factors relates to other previously identified factor structures within the domain of interest.

Factor Analysis: Procedures

The following is a summary of the procedures used in the factor analysis for each scale.

For the YITS data, consideration was given to the effect of the language of interview on responses. Any changes in the presentation of items may introduce substantial bias in responses. (Fowler, p. 74, 1995). With the case of languages, this bias may affect the responses due to imprecise translations, creating potentially different interpretations of the question. This, in turn, may lead to different response patterns in those respondents who were administered the questionnaire in English and those who were administered the questionnaire in French.

To examine the possibility of translation bias, the data was first divided into two samples according to the languages of questionnaire administration: English and French. Comparison of the results from each linguistic half-sample was undertaken. Greater dissimilarities between the results could indicate a greater translation bias.

The data from both the English and the French samples were further randomly split into two half-samples. The first half-sample for both the French and the English sample was considered as a test sample and the second half-sample as a verification sample. The verification sample was used to independently confirm the structure identified in the first half-sample.

A principal component based factor analysis was carried out separately on each linguistic sub-sample to determine how many factors should be extracted from the data. Factor loadings of each variable were compared across the half-samples and across language. Loadings were examined under a rotation of the factor loading matrix when the concept had two or more factors associated with it. A rotation involves a shift in the coordinate axis of the loadings matrix. When it is not easy to interpret a loading, a more readily interpretable solution may be obtained from a rotation. Under a rotation, one would hope to obtain a pattern of factor loading such that a variable loads highly on one factor and has only low to moderate loadings on the other factor or factors.

In this factor analysis, the items for each respondent were multiplied by the respondent's normalized weight. A normalized weight was calculated by dividing a respondent's survey weight (see **section 10.2** for a discussion on weighting in YITS) by the average weight of all eligible respondents. Thus, in theory, the sum of the normalized weights is equal to the sample size of the eligible respondents. These weights were used in forming a design consistent estimate of the variance-covariance matrix.

Within a questionnaire, some questions are positively oriented such as *I got along well with teachers*. Other questions are negatively oriented such as *Learning in class was useless*. In order for the effects on a scale of negatively orientated items not to cancel out the effects of positively oriented items, and thus cause the scale not to perform properly, the negative items were reoriented to make them positive. Letting **k** be the number of categories for a particular item, a positively oriented item is obtained from a negatively oriented one by subtracting the value of the item from **k** + 1. ((**k**+1)- item value). On a four point scale, if a respondent indicated a value of 2 on a negatively oriented question, then the positively oriented response value would be 3 (3 = 5-2). This reorientation has to be done before the score is calculated in order to properly estimate the scale item internal consistency (Cronbach's alpha) and to estimate scale scores. (See **section 4.3.2.3** for a discussion of Cronbach's alpha).

It should be noted that in addition to the estimated scores, the items for each scale are included on the micro data file. This will allow researchers to consider alternate factor structures. The original values, before any reversal of values took place, have been retained for all these items.

4.3.2.2 Estimation of the scores

The results from the factor analysis were used to determine what items loaded onto each factor (**section 4.3.2.1**). Once the factor structures were analyzed and the items to be included in each factor were verified, the next step was to estimate the respondent's latent construct score for each factor. Two approaches were investigated, a linear composition based upon the linear factor loadings (the standard classical method) and non-linear maximum likelihood estimation based upon IRT.

Factor analysis requires that the data have the property of interval or ratio data; this property assumes that the distance between each neighbouring answer category of the question is the same. However, the data only has the property of ordinal data. In Likert-type scales with items where the answer choices are: 1 (strongly disagree), 2 (disagree), 3 (agree), 4 (strongly agree), it is assumed that the distance between strongly disagree and disagree is the same as that between agree and strongly agree in the respondent's perception.

IRT is able to control better for the categorical nature of the data than factor analysis. In addition, other countries participating in similar studies such as the *Programme for International Student Assessment* (PISA) favoured using an IRT response (Wolfram, 2002). The software program PARSCALE (Muraki & Bock, 1997) was applied to calculate the IRT scores and the estimates of the measurement errors.

The IRT scores and their respective standard errors were estimated using weighted maximum likelihood (see Warm, 1989) and applying a generalized partial credit model. The generalized partial credit model is an extension of the two parameter logistic distribution to polytomous data. For estimating IRT scores, the population distribution of the scores was specified to have a mean of 0 and a standard deviation of 1. A response pattern of mostly extreme positives (i.e., strongly agree to all positive questions and strongly disagree to all negative questions) is most likely to have been produced by an individual with a highly positive standard score. Conversely, a response of mostly extreme negative values is most likely to have been produced by an individual with a highly negative score. A typical mix of responses (few extreme responses) is likely produced by an individual with a score of around 0.

4.3.2.3 Scale reliability

Reliability, when discussing scales, refers to the accuracy, dependability, consistency or repeatability of score results. More technically, it refers to the degree to which the scores are free from measurement errors. It can be interpreted as a holistic measure of the accuracy of a scale, in that it describes the proportion of the population variance in

scores that can be explained by the population variance in the underlying construct. Two measures of reliability are commonly used in examining scales, Cronbach's alpha and the Index of Reliability.

Cronbach's Alpha and its Interpretation

Cronbach's alpha is a measure of the internal consistency of the items within a factor. It is based upon the average covariance of items within the factor and assumes that the items within a factor are positively correlated with each other; they attempt to measure a common entity or construct.

Cronbach's alpha has several interpretations. It is theoretically related to the correlation between the scale being used and all other possible scales containing the same number of items that could be constructed from a hypothetical universe of items that measure the same characteristic of interest. With this measure, the analyst is able to obtain the expected relationship between the scale that was used and all other possible scales that measure the same concept. Since Cronbach's alpha can be interpreted as a coefficient of determination its values range from 0 to 1.

Index of Reliability

While Cronbach's alpha estimates the reliability as the internal consistency of a scale's items, a more accurate estimate of reliability is the index of reliability, η which incorporates the IRT characteristics of each item. Similar to Cronbach's alpha, values of this index closer to 1 indicate a greater accuracy and denote better measurement properties of the scale (Crocker and Algina, p. 352, 1986).

$$\eta_{J.\theta} = \sqrt{1 - \frac{\sigma^2{}_E}{\sigma^2{}_J}} \quad (4.3.1)$$

Figure 4.3.1 Index of Reliability

For a given scale σ^2_E is the weighted average measurement variance across the sample, and σ^2_J is the estimated variance of all scores in scale j. Although the value of η_j obtained will be similar in magnitude to that of Cronbach's alpha, it is a more accurate measure of the reliability of the final scores that have been produced.

4.3.2.4 Scale validity

In order to assess whether the estimated scale scores behave according to the theoretically proposed conception of the model, validity tests were performed on the scales. These tests involve evaluating the proposed scales or sub-scales by comparing their estimated scores on the scales to scores on other relevant scales or to the values of other relevant variables. They can also involve the comparison of different identifiable groups of respondents on the scale of interest. Testing for scale validity involves knowledge of the subject matter involved in the analysis and in particular, which variables or scales are expected to be related to the scale of interest and the form that this relationship is expected to take.

4.3.3 The high school engagement scale

4.3.3.1 Why is there interest in a school engagement model?

One of the most prevalent problems for educators today is the emotional, intellectual and physical withdrawal of students from formal schooling (Voekl, 1996). A student's engagement, or involvement with school, has been linked to academic achievement

and to the probability of graduating from high school. In addition, "as children grow, this social, emotional and intellectual withdrawal, when accompanied by low grades, is associated with several serious problems including truancy, juvenile delinquency…" (Voekl, 1996) Research indicates that children and youth who are engaged with school are likely to achieve success and become school graduates (Voelkl, 1997, 1995, cited in Norris et al, 1998).

Ensuring that Canadian youth remain in and succeed at school is a high priority for many groups within Canada including parents, educators, other youth workers and different levels of government. To fully compete in national and international labour markets and actively participate as citizenry, youth need to complete their education and to achieve high academic standards in school. "Youth who have the needed skills and knowledge are better equipped to succeed as global trade expands and economic opportunities open …. [whereas] youth who have dropped out of high school … do not yet possess the skills and knowledge to effectively compete in the labour market." (Statistics Canada, 2002) As school engagement plays a large role in keeping youth in school, and remaining in school helps youth to meet their full potential as citizens and as workers in the new economy, it is important to identify and try to understand the factors that influence a youth's engagement in school.

4.3.3.2 What is the YITS high school engagement scale?

To obtain information about school engagement, the YITS team designed the High School Engagement scale. While school engagement is an important issue throughout a student's entire school career (Voekl, 1996), the High School Engagement Scale administered to the YITS 18-20 year old cohort only measured engagement for the student's last year of formal schooling. The engagement questions focussed on this last year since the memory of feelings related to that year would be the strongest. Moreover, students have likely changed schools several times throughout their academic career. These students might have had very different school-related experiences at different institutions and even in different academic years at the same institution. Since the type of experiences that a student receives may differ depending upon the year of focus, the student's response to the items may also be different depending upon the questionnaire's timeframe.

The last year of formal education was usually high school (approximately 99% of respondents). However, this last experience could have been in either junior high or elementary school if the respondent had left formal schooling before starting high school. In these cases, the information collected from the respondent regarding school engagement relates to that last year of schooling. Although the High School Engagement scale asks youth to focus upon their last year of formal education (high school, junior high or elementary school), this last year need not have been taken in either Canada or in the United States. Indeed, as long as some of their education, either at the elementary or at the secondary level, was taken in either Canada or in the United States the scale-related questions were administered to them.

The High School Engagement scale is composed of the Academic Engagement and Social Engagement sub-scales. In developing a scale for a concept such as school engagement, the items to test and the scale structure are usually created based upon a theoretical model and /or previous research.

For the PISA/YITS Reading cohort and for the upcoming PISA/YITS Mathematics cohort, an adapted version of items developed by Jeremy D. Finn (Department of Counselling and Educational Psychology, State University of New York at Buffalo) for assessing the academic participation of young adolescents is included (Finn, 1993). Also included in the scale are selected items from Kristin Voelkl's Identification with School Scale that covers academic and social identification. She refers to these concepts as school identification (Voelkl, 1996). To these scales, a few additional

items were added[7] to more adequately represent the concepts proposed in the theoretical model. Using these items, the Youth in Transition Survey adapted the model and definitions of school engagement provided in Norris et al (1998). Administration constraints due in part to the fact that the potential content for the YITS 18-20 year old cohort questionnaire was more broadly defined than that of the PISA/YITS Reading cohort allowed for only a subset of the items from the PISA/YITS Reading cohort questionnaire to be included in the 18-20 year old cohort questionnaire. This was done in an attempt to decrease respondent burden. (Statistics Canada, 2000)

[7]Valuable advice regarding the items chosen for assessing school engagement and related concepts was also given by Peter Dietsche, Director of Humber Research Network, Humber College, Ontario; the late Sid Gilbert, Professor, Department of Sociology and Anthropology, University of Guelph, Ontario; and Herb O'Heron, association of Universities and Colleges of Canada.

4.3.3.3 Structural model of high school engagement

A theoretical construction of the high school engagement model follows below.

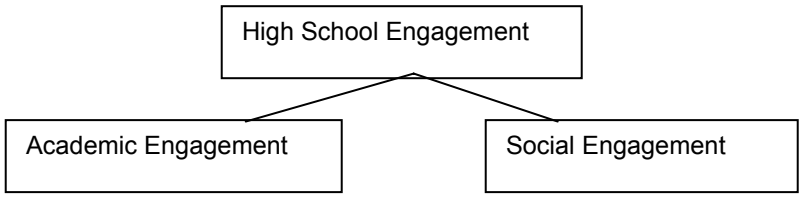


Figure 4.3.2 Model of High School Engagement for the YITS 18-20 year old Cohort

Academic Engagement

Academic Engagement is defined as the identification with and behavioural involvement (participation) in the academic aspects of school. Academic aspects of school include the students' dealings with teachers, curricula, and the school governance. Students who do not identify with school may be predisposed to a pattern of negative school behaviour and to eventual withdrawal from school. By contrast, students who do identify with school are also expected to have a higher academic achievement.

Social Engagement

Social Engagement is defined as the identification with and behavioural involvement in the social aspects of school (the school social life). It involves both a feeling of belonging to the school's social environment and a sense of fit between the individual and the school. Social aspects are informal, out-of-classroom interests and activities associated with the school. These include students' relationships with peers, their extracurricular activities and their contact with teachers outside of the classroom.

4.3.4 Analysis of high school engagement scale

The main steps in the development of the scales presented in **section 4.3.2** were applied to the high school engagement model illustrated in **section 4.3.3.3.** Analysis of the YITS 18-20 year old cohort data revealed that the data supported two factors: Academic Engagement and Social Engagement. Nine items loaded on the Academic Engagement sub-scale and four items loaded on the Social Engagement sub-scale.

The empirical structure of the High School Engagement Scale for the PISA/YITS Reading cohort (see the PISA/YITS Reading cohort User's Guide for more information) and the empirical structure of the 18-20 year old cohort's High School Engagement scale are different. This difference may be due to either increased measurement error or to construct bias because of the limited number of items selected for the YITS 18-20 year old cohort questionnaire. Researchers are cautioned to be aware of this fact when comparing results between the two High School Engagement scale models.

4.3.4.1 Testing model validity

The fifteen items that were hypothesized to define school engagement were administered to a sample of 22,310 eligible 18-20 year olds from the respondents in cycle 1 of YITS. (See **section 4.3.3.2** for a discussion of eligibility for YITS scale questions). This sample was split by language of questionnaire administration: English and French. Both the English and French sub-samples were then further split into two random samples and a factor analysis procedure was implemented to determine which factors were supported by the data.

The results across sub-samples and across languages were similar. This was an indication of the scale remaining stable across different samples and populations. Hence the analysis was repeated using all eligible respondents. This resulted in a two-factor model – Academic Engagement (YSHACES1) and Social Engagement (YSHSOES1) – having 9 items, and 4 items respectively. Two items were dropped from the analysis because they did not load strongly on either of the two factors (F7A) or because the item loaded differently on the English and the French sub-samples (F25).

The items retained for each factor, their description and their factor loadings are provided below. The name of the item in the questionnaire booklet is not the same as the name provided in the YITS codebook. For easy reference, both names are provided in **Tables 4.3.2** and **4.3.3**.

Table 4.3.2 Academic Engagement Items Description and Loadings

Item Code Questionnaire	Item code Codebook	Item Description	**Factor Loadings**
F6	F6	Hours in a week on homework	0.52
F11	Fc11_F11	I got along well with teachers	0.59
F16	Fc11_F16	I just wanted to get by	0.65
F17	Fc11_F17	I paid attention to the teacher	0.70
F18	Fc11_F18	I was interested in what I was learning	0.65
F20	Fc11_F20	I completed my homework on time	0.67
F21	Fc21_F21	Learning in class was useless	0.47
F28	Fc21_F28	School often was a waste of time	0.52
F30	Fc30_F30	Number of times skipped classes in a month	0.60

Table 4.3.3 Social Engagement Items Description and Loadings

Item Code Questionnaire	Item Code Codebook	Item Description	**Factor Loadings**
F19	Fc11_F19	I felt like an outsider	0.50
F22	Fc21_F22	I was treated with as much respect as others in my class	0.54
F24	Fc21_F24	I had friends at school I could talk to	0.63
F29	Fc21_F29	People at school were interested in what I had to say	0.64

The factor loading for the items on both sub-scales were similar in magnitude. According to the Comrey and Lee benchmark for rating scale loadings (**section 4.3.2.1**), the loadings for the Academic Engagement sub-scale with values from 0.48 to 0.70, range from fair to excellent, and the loadings for the Social Engagement sub-scale, with values from 0.50 to 0.64, range from fair to very good.

4.3.4.2 Estimating scores

Scores on the two subscales were estimated using item response theory by the program PARSCALE (Muraki & Bock, 1997). An estimated score was able to be obtained for most respondents. By taking a simple average of the scores of the two sub-scale variables, an overall measure of high school engagement (YSHSCES1)

was derived. The measurement standard error of the scores for each respondent on both of the sub-scales and the overall school engagement scale was estimated and is included along with the scores on the YITS dataset.

Due to a few respondents not answering any of the questions upon which the subscale was based**,** a number of scores were unobtainable. (See **section 7.4** for a discussion of the coding of missing values in YITS). Fourteen cases for the Academic Engagement sub-scale and forty-nine cases for Social Engagement sub-scale were not derived for this reason. When either the Academic sub-scale or the Social Engagement scores were missing, the high school engagement scale score was set to missing. Including the estimated score from only one of the two components that made up the High School Engagement scale would bias the results of secondary analyses. For these students, High School Engagement would be wrongly estimated as either Social or Academic Engagement, neither of which is conceptually interchangeable with overall school engagement. Forty-nine cases were assigned missing scores for overall school engagement. The much larger number of unestimated scores for the Social Engagement sub-scale may be due to the fact that there were fewer items that loaded on this scale and thus less information was available from which to estimate the scores.

4.3.4.3 Testing scale reliability

Two common measures of reliability, Cronbach's alpha and the index of reliability were estimated. The value of Cronbach's alpha for the items in the Academic Engagement sub-scale is 0.80 and is 0.61 for the items in the Social Engagement sub-scale. A value of the index of reliability of 0.93 is obtained for Academic Engagement, 0.63 for Social Engagement and 0.85 for the overall School Engagement. Both sets of indices suggest a greater amount of measurement error in the Social Engagement sub-scale. This greater imprecision is likely due to the smaller number of items in the Social Engagement sub-scale. However, researchers should use these reliability estimates and the standard errors of the scores provided with the micro data file to determine whether or not these sub-scales are reliable enough for their purposes.

4.3.4.4 Testing scale validity

Below is a list of some of the variables that were found to be correlated with the Academic Engagement sub-scale and/or with the Social Engagement sub-scale. The direction of the correlation and the strength of the relationship between the variable and the sub-scale are also provided. These correlations provide further evidence of the validity of the scales.

Table 4.3.4 Academic Engagement Correlations

Item Code Questionnaire	Item Code Codebook	Item Description	Correlation Strength	Correlation Direction	Hypothesis
C0A	C0A	Overall grade	High	Positive	Youth who get higher grades are more likely to be academically engaged in school
C3	C3	Math grade	Medium	Positive	See above
C8	C8	Main Language grade	High	Positive	See above
F10	FC10_F10	Number of friends planning further Education	Medium	Positive	Youth might associate with other friends who have similar interests to their own. Those youth who have many friends who are planning to take further education may be more academically and socially motivated than

Item Code Questionnaire	Item Code Codebook	Item Description	Correlation Strength	Correlation Direction	Hypothesis
					those who have a few or no friends who are planning further education
F31A	FC30F31A	Number of times/month drinking	High	Negative	Youth who tend to drink a lot more when they do drink and who drink frequently might tend to be less academically engaged
F31B	FC30F31B	Number of drinks when drinking	High	Negative	See above
F32	FC30_F32	Number of times using hashish or marijuana	High	Negative	Deviant behaviour patterns (such as drug use) may be an indication of students who are less academically engaged in school
---- (derived variable)	HSLD03	High school status	High	Positive	HS continuers and graduates should show higher levels of academic and Social Engagement than HS leavers
MC3	MC3	Highest Level of education you think you will get	High	Positive	Youth whose educational aspirations are higher are more likely to be academically engaged
V24	V24	Importance of more education after high school	Medium	Positive	Parents may influence their child to also value education and thus have higher Academic Engagement

Table 4.3.5 Social Engagement Correlations

Item Code Questionnaire	Item Code Codebook	Item Description	Correlation Strength	Correlation Direction	Hypothesis
C0A	C0A	Overall grade	Medium	Positive	Youth who are socially engaged in school may enjoy school more and tend to be more active learners. On the other hand, some youth are too socially involved in school and this would interfere with their overall grade.
C3	C3	Main Math grade	Medium	Positive	See above
C8	C8	Main Language grade	Medium	Positive	See above
F10	FC10_F10	Number of friends planning further Education	Medium	Positive	Youth might associate with other friends who have similar interest to their own. Those youth who have many friends who are planning to take further education may be more academically and socially motivated than those who have a few or no friends who are planning further education.
---- (derived variable)	HSLD03	High school status	Medium	Positive	HS continuers and graduates should show higher levels of Academic Engagement and Social Engagement than HS leavers.
MC3	MC3	Highest level of education you think you will get	Medium	Positive	Youth who enjoy school will have higher educational aspirations.

Most of the items listed in **Table 4.3.3** had a relatively high correlation with the Academic Engagement sub-scale whereas those listed in **Table 4.3.4** had a medium correlation with the Social Engagement sub-scale. This lower average correlation with items is likely due to the greater measurement error of the Social Engagement sub-scales. Possible reasons for this were described above (**section 4.3.4.3**). Increasing the error variance of a variable has the effect of reducing the magnitude of the observed correlations between it and other variables, an effect known as *attenuation*. (Glass and Hopkins, 1996, pp. 121- 126).

4.3.5 Scale references

Analysis of Scales for YITS/PISA Cycle 1 Main Survey, Statistics Canada.To be available fall of 2003

Bock, R.D, and Muraki, E. (1997), PARSCALE - *IRT: Item Analysis and Test Scoring for Ratings-Scale Data.*Chicago, Illinois. Scientific Software International.

Bowley, J.W. and McMullen, K. (2002), *At a Crossroads: First Results for 18 to 20-Year-old Cohort of the Youth in Transition Survey* Catalogue no. RH64-12 / 2002E. Statistics Canada

Crocker, L. and Algina, J. (1986), *Introduction to Classical and Modern Test Theory.* Belmont, California.Wadsworth Group

Fidell, L.S. and Tabachnick, B.G. (1996), *Using Multivariate Statistics.* Northridge California. Harper Collins College Publishers

Finn, J.D. (Ed.), (1993), *School Engagement and Students at Risk*, Washington, DC, National Centre for Education Statistics.

Fowler, F.J. (1995), *Survey Research Methods: second edition.* London: Sage Publications

Gorsuch, R.L. (1997), "Exploratory Factor Analysis: Its Role in Item Analysis,". *Journal of Personality Assessment,* vol. 68, 532-560

Glass, G. and Hopkins, K.(1996), *Statistical Methods in Education and Psychology 3rd ed.* Boston: Allyn and Bacon.

Likert, R.(1932), "A Technique for the Measurement of Attitudes," *Archives of Psychology*, No.140, 1-55.

NLSCY Scales. Internal Document Statistics Canada

Norris, C., Pignal J. and Lipps G., (1998), *Measuring School Engagement in the National Longitudinal Survey of Children and Youth: Using Factor Analysis as a Tool in Content Development*, Internal Document. Statistics Canada.

Schulz, W. (2002), "Constructing and Validating the Questionnaire Indices," In Adams, R., M.Wu (Eds.), *PISA 2000 Technical Report* (pp.217-252), Paris: Organisation for Economic Co-operation and Development.

Voelkl, K. E. (1996), "Measuring Students' Identification with School," *Educational and Psychological Measurements* , vol. 56, 760-770.

Warm, T. (1989), "Weighted Likelihood Estimation of Ability in Item Response Theory," *Psychometrika,* vol. 54, 427-450

Statistics Canada.(2000), *Youth in Transition Survey – Project Overview* .
Catalogue no. MP32-30/ 00 – 5E. Statistics Canada

5.0 Survey Methodology

Definition of the YITS population

The YITS target population for the 18-20 year-old cohort comprises residents of the ten provinces of Canada who were born in the years 1979 to 1981. A large portion of the questionnaire for Cycle 1, which was administered from January to April 2000, is devoted to profiling these individuals' education and labour market activities during the reference year of 1999, when they were 18 to 20 years old.

The sample design of the YITS 18-20 cohort is determined to a large degree by the sample design of the Labour Force Survey (LFS). This applies to the stratification, allocation and selection of the sample as well as the estimation methods.8

5.1 LFS Sample Design

The LFS is a monthly survey that collects labour market data from a national sample of about 60,000 dwellings. The current design was implemented at the end of 1994 following a redesign program that included a reassessment of the survey's principal role as a provider of current labour market information as well as a central vehicle for conducting household surveys within Statistics Canada.

5.2 LFS Target population

The LFS sample is representative of the civilian, non-institutionalised population 15 years of age or older in Canada's ten provinces. Specifically excluded from the survey's coverage are residents of the Yukon, Nunavut and Northwest Territories, persons living on Indian Reserves, full-time members of the Canadian Armed Forces and inmates of institutions. These groups together represent an exclusion of approximately 2% of the population aged 15 or over.9

5.3 LFS Stratification

The LFS sample is based upon a stratified, multistage design employing probability sampling at all stages of the design. The design principles of the LFS are the same for each province.

Primary strata

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

The intersections of the two types of regions form primary strata for the LFS. Note that a third set of regions, Census Metropolitan Areas (CMAs), is also respected by stratification in the current LFS design, since each CMA is also an EIER.

8 For a detailed account of the LFS methodology, see Gambino, J.G., Singh, M.P., Dufour, J., Kennedy, B. and Lindeyer, J. (1998), *Methodology of the Canadian Labour Force Survey*, Statistics Canada, Catalogue no. 71-526.

9 Since 1992, the LFS has been administered in the Yukon, using an alternative methodology that accommodates some of the operational difficulties inherent to remote locales. To improve reliability due to small sample size, estimates are available on a three-month average basis only. In 2000, the survey was extended to the Northwest Territories and Nunavut. No estimates for any of the territories are included in national totals.

Types of areas

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

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

Approximately 1% of the LFS population is found in remote areas of provinces that are less accessible to LFS interviewers than other areas. For administrative purposes, this portion of the population is sampled separately through the remote area frame. Places with fewer than 10 households or 25 persons and Census Enumeration Areas (EA) with fewer than 25 households are omitted from the design.

Secondary strata

In urban areas with sufficiently large numbers of apartment buildings, the strata are grouped according to those based on apartment frames and those based on area frames. The apartment list frame is a register that is based upon information supplied by CMHC and is maintained in the 18 largest cities across Canada. The purpose of this frame is to ensure better representation of apartment dwellers in the sample as well as to minimise the effect of cluster growth due to construction of new apartment buildings. In the major cities, the apartment strata are further stratified into low-income strata and regular strata.

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

In rural areas, the population density can vary greatly from relatively high population density areas to low population density areas, resulting in the formation of strata that reflect these variations. The different stratification strategies for rural areas were based not only on concentration of population, but also on cost-efficiency and interviewer constraints. Also, within each of the secondary strata in rural areas further stratification is carried out where necessary to reflect differences among a number of socio-economic characteristics within each stratum.

The remote area frame is stratified only by province.

5.4 LFS Cluster delineation and selection

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

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

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

Other procedures are applied in rural and remote areas. Within each rural stratum, six EAs or two or three groups of EAs are sampled as clusters, whereas remote settlements within each province are sampled proportional to the number of dwellings in the settlement.

5.5 LFS Dwelling selection

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

5.6 LFS Sample rotation

The LFS employs a panel design whereby the entire monthly sample of dwellings consists of six panels or rotation groups, of approximately equal size. Each of these panels can be considered to be representative of the entire LFS population. Dwellings are in the LFS for six consecutive months. Each month a new panel of dwellings selected from the same or similar clusters replaces the sample dwellings in one of the rotation groups.

This rotation pattern has the statistical advantage of providing a common sample base for month-to-month comparisons of LFS characteristics. It also ensures that the sample of dwellings constantly reflects changes in the current housing stock and helps to minimise the respondent burden and non-response that could result if households were to remain in the sample longer than six months. Surveys that use the LFS frame or sample can take advantage of the rotation group feature to use larger or smaller sample sizes than that of the LFS.

5.7 Household members eligible for the LFS

The first month a dwelling is in the LFS a roster containing information on the household composition is completed. Demographic information including the name, sex, date of birth and education level is obtained for all persons for whom the selected dwelling is the usual place of residence. LFS information is obtained for all civilian household members 15 years of age or older.

When the dwelling is contacted in subsequent months the roster is updated to reflect changes in household membership from the previous month. Response burden is minimised for the elderly (70 years of age or older) by carrying forward their responses for the initial interview to the subsequent five months in the survey.

5.8 YITS Target population and sample design

As noted at the beginning of the chapter, the target population for the 18-20 cohort of the YITS comprises residents of the ten provinces of Canada who were born in 1979 to 1981. These individuals turned 18 to 20 during 1999, the reference year for cycle 1. The LFS exclusion of full-time members of the armed forces and persons living on Indian reserves or in northern and remote areas also applies to the YITS.

The design implemented for the 18-20 cohort of the YITS is based on certain groups of households that were in the LFS between January 1997 and December 1999. A sample of 29,950 individuals was selected for Cycle 1 from a combination of households in 30 LFS groups that rotated out from January 1997 to November 1999 and households from all six rotation groups in the December 1999 LFS.

To select the YITS cohort 18-20 sample, the set of in-scope households within each of the 36 groups was identified. In-scope households were those with at least one household member with a high probability of being in the YITS target population. The majority of individuals were identified by year of birth - for household members without a year of birth on the LFS roster, the age and the month of the LFS were used to calculate an estimated age as of December 31, 1999. Thus households were included in the initial YITS sample if at least one household member had

- a specified year of birth from 1979 to 1981, or
- an unspecified year of birth and an estimated age of 17 to 21 years.10

From the initial sample of 29,950 households, 786 were treated as a priori non-response and were excluded from the final sample selection stage for the YITS. Two-thirds of these households were eliminated from the YITS sample because they already had a household member in the longitudinal sample for the National Longitudinal Survey of Children and Youth and/or the National Population Health Survey. Most of the other excluded households were eliminated because they had refused forthcoming LFS interviews or their LFS interviews were conducted in person rather than by telephone.

The final stage of sampling for the YITS was the selection of one household member in the YITS target population from each of the remaining 29,164 in-scope households. In the minority of households that did have more than one person in the YITS target population, one was selected with equal-probability systematic sampling.

10 An age interval of 17 to 20 was applied for the rotate-out group from January 1997 to avoid flooding the sample file with a large number of cases that had a high probability of being out of scope. This was the last rotate-out group based on the old LFS questionnaire that collected age rather than date of birth. The actual rotate-out month combined with the year-of-birth definition of the YITS target population resulted in a very low probability of a person with estimated age 21 actually being born in 1979 to 1981 so these persons were classified as not eligible for the YITS sample selection.

5.9 YITS Sample Size By Province

The following table shows the sample sizes by province

Province	**Sample Size**
Newfoundland and Labrador	1,411
Prince Edward Island	780
Nova Scotia	1,826
New Brunswick	1,715
Quebec	5,881
Ontario	8,520
Manitoba	1,952
Saskatchewan	2,105
Alberta	2,380
British Columbia	2,594
All provinces	**29,164**

6.0 Data Collection

Data collection for Cycle 1 of the YITS took place between January and April 2000.

6.1 The Questionnaire

The YITS questionnaire for the 18-20 year-old was developed using CASES software for a computer assisted telephone interview (CATI). Some advantages of CATI are: question flows are controlled by the computer which allows for a more complex interview with little added burden to the interviewer; data capture occurs during the interview with on-line edits verifying that the data have been captured properly; and the system is able to take care of scheduling appointments, accounting for time zones, etc.

6.2 Training

Given the large survey sample size, it was decided to conduct the YITS from five Statistics Canada regional offices: Vancouver, Winnipeg, Montreal, Sturgeon Falls and Halifax. A large number of interviewers and senior interviewers were required to work on the survey so the team implemented a two-phase classroom training plan. Prior to the classroom training, the participants received their self-study materials, which included interviewer and content manuals.

The first phase of classroom training took place in Head Office in Ottawa, where up to two senior interviewers and one project manager from each regional office were invited to a four day training session.

The senior interviewers and project managers who were trained during the first phase returned to their regional offices where the second phase of classroom training took place.

Once the second phase of training was completed, interviewing began. The survey team members who had attended the training sessions stayed in the regional offices for several days afterwards, to observe the start-up of the survey. This was beneficial to everyone as it allowed the survey team to see first-hand the survey in the field while giving the regional offices someone on-site who could help answer survey-specific questions.

6.3 Supervision and Control

All interviewers are under the supervision of a staff of senior interviewers who are responsible for ensuring that interviewers are familiar with survey concepts and procedures, periodically monitoring their interviewers and reviewing their work. Senior interviewers ensured that prompt follow-up action was taken for refusals and other non-response cases. The senior interviewers are, in turn, under the supervision of the Regional Office project managers.

6.4 The Interview

As noted in Chapter 5, a single respondent had been identified as a potential respondent for the YITS. In order to facilitate contacting the selected respondent, the sample file included the respondent's name, address and telephone number, as well as additional addresses and phone numbers where possible. This provided the interviewer additional "leads" to find the respondent, should attempts with the original telephone number prove unsuccessful.

If the selected respondent could not be located, then the interview could not take place. Proxy reporting was not permitted.

By the end of data collection, the total average interview time per case/unit (including contact, tracing, and call-backs) was 46 minutes. The total average questionnaire time (Modules A through Z) was 34 minutes.

6.5 Follow-up of Non-response

Interviewers were instructed to make all reasonable attempts to obtain the YITS interviews with the selected respondent. For cases in which the timing of the interviewer's call was inconvenient, an appointment was arranged to call back at a more convenient time. For cases in which there was no response, there was follow-up.

Non-response can be viewed from a number of perspectives. In the context of the YITS, there are 3 types of non-response:

1. Refusals

If a respondent refused to participate, then the case was coded as a "refusal". An interviewer specifically responsible for refusal conversions would then access the case and call back the respondent to attempt to persuade him/her to participate.

2. Partial interviews

If a respondent failed to complete the interview (completion meant the respondent had answered the data sharing question - this was the 2nd last question before respondents were asked if they wished to provide any general comments) the case was assigned a "partial" code. Again, an interviewer would call back and attempt to persuade the respondent to complete the interview.

3. Non data-sharers

The data-sharing question at the end of the questionnaire asked if the respondent was willing to share their answers with the survey's client, Human Resources Development Canada. If a respondent refused to share, then the case was coded "non-sharer" and an interviewer would then re-contact the respondent and attempt to persuade him/her to share his/her data.

7.0 Data Processing

The main output of the YITS is a "clean" microdata file. This section presents a brief summary of the processing steps involved in producing this file.

7.1 Data capture

As mentioned previously, responses to the questionnaire were captured directly by the interviewers at the time of collection, using computer-assisted telephone interviewing. A partially edited file was transmitted to Ottawa for further post-capture processing.

7.2 CAI Editing

CAI editing occurred in the Regional Offices during data collection. Various edits and checks were built into the questionnaire to check for data paths and flows, internal consistency and to ensure the quality of the information collected.

All flow patterns were built into the system. Therefore, interviewers could only access questions that logically followed from the respondent's previous answer.

The first types of error treated were consistency errors. Consistency edits were most frequently used with dates. Whenever a consistency edit was triggered, interviewers had the opportunity to change the response that had been entered in error.

A second types of error treated were range errors. Range checks were used for continuous variables. In some cases the system instructed the interviewer to either ask the respondent to confirm a value or to enter a new value should the value have fallen outside the 'normal' range.

7.3 Minimum completion requirements

For all surveys it is essential that a minimum number of key fields are completed. In the case of the YITS, records were retained so long as high school status could be derived (see ***Section 4.2.1***).

7.4 Head Office editing

The first stage of survey processing undertaken at Head Office was the "pre-edit".

- The purpose of the pre-edit was to carry out some basic formatting and preliminary editing to make further processing easier

Once the data had been pre-edited, some analysis/editing were done to look at:
- post-secondary institutions/programs that had been made ineligible during collection
- duplicate post-secondary institutions/programs
- jobs that had been made ineligible during collection

Sometimes the interviewer entered an answer in the "Other specify" field even though the answer appeared in the response list. All "Other specify" text fields were reviewed, and if possible, the responses were re-coded back into the existing response lists. Of note are two questions in the questionnaire where the text fields could not be re-coded: L1H (other source of post-secondary financing) and Q1_Q1H (other types of volunteer activity).

Once the data had been returned to Head Office, the following variables were coded and the new codes were added to the data file:

- Major field of study

  The variables that contained the text fields for major field of study were sent for automated coding. The coding system was an exact match system that matched the YITS text against the hundreds of thousands major field of study examples collected by the Census. There was an 80% exact match rate. The remaining 20% of text fields were manually coded. All codes were then added to the data file.

- Industry and Occupation variables

  The variables that contained the text fields for industry and occupation were sent to Operations and Integrated Division for manual coding. The coding produced SOC and NAICS codes for all jobs collected in the YITS and these codes were added to the data file.

Institution/campus codes were verified in Head Office and were evaluated at being 85% correct. The 15% of records in error were corrected and the revised codes were added to the data file.

Coding structure for YITS

Researchers are cautioned when interpreting data from numerical variables to be aware of valid skips or not stated codes potentially affecting their estimates. Two possible solutions for handling these codes in numerical variables include excluding records with these codes or imputing values for them.

- *Assignment of Valid Skip codes*

The flow of the questionnaire during the interview was controlled by the computer system. If an interviewer started down a path of questions and then realised that it was not the proper path for the respondent, it was possible for the interviewer to return to a previously answered question, change the answer, and take another path. When this happened, however, the system retained the answers in both paths. Once the data were returned to Head Office it was important to determine the proper path so that the correct data were preserved and "valid skip" codes were assigned to the paths not taken.

Valid skips are assigned using a top-down editing approach. In other words, if a respondent answers "yes" to Q1 and a yes answer brings him/her directly to Q10, then questions Q2 through to Q9 for this respondent are set to valid skips.

The code that is used to identify a valid skip is 6. For variable lengths greater than 1 byte, 9's are added before the 6. If the variable has no decimals the values would be 96, 996, 9996, 99996, etc. For the variables with decimals (all such variables that are assigned valid skips codes have two decimals) the values would be 9.96, 99.96, 999.96, etc.

- *Assignment of Not stated codes*

Not stated codes are assigned to a question where ever a respondent should have provided an answer but for some reason did not provide one. An exception to this rule occurred with the scale scores and their corresponding standard errors. Respondents who did not receive the scale items (see section 4.3.3.2 for eligibility to receive scale items) or who did not answer any of the questions related to either the Academic Engagement or the Social Engagement sub-scales were assigned a code of not stated for that sub-scale. Respondents who had one of these sub-scales set to not stated were also assigned a code of not-stated for the over-all High School Engagement scale.

The code that is used to identify "not stated" is "9". For variable lengths greater than 1 byte, 9s are added before the 9. If the variable does not have any decimals, the values would be 99; 999; 9999; 99999, etc. For most variables with decimals (with the exception of scales, all such variables that are assigned "not stated" codes have two decimals) the values would be 9.96; 99.96; 999.96, etc. If the variable is a scale score it would be 99.99999, and if it corresponds to a scale standard error it would be 9.999999.

- *Assignment of Not applicable codes*

Not applicable codes have been used only for certain derived variables (see ***section 4.2.3***).

The code that has been used to identify not applicable is 5. For variable lengths greater than 1 byte, 9's are added before the 5 i.e. (as all variables that are assigned Not applicable codes have no decimals) 95, 995, 9995, 99995, etc.

- *Don't know and Refusal codes*

For most questions in the YITS, respondents were able to answer "I don't know" or "I refuse". During the pre-edit, these answers were changed to 7's and 8's. For variable lengths greater than 1 byte, 9's are added If the variable has no decimals the values would be 97, 997, 9997, 99997, etc, and 98, 998, 9998, 99998, etc. For the variables with decimals (all such variables that are assigned don't know and refusal codes have two decimals) the values would be 9.97, 99.97, 999.97, etc, and 9.98, 99.98, 999.98, etc.

7.5 Creation of Derived Variables

A number of data items on the microdata file have been derived by combining items on the questionnaire in order to facilitate data analysis. A complete list of derived variables is provided in Section 4.2.

In general, a derived variable was not calculated if any part of the equation was not answered (e.g., don't know, refusal and not stated). In these cases, the code assigned to the derived variable was "not stated".

7.6 Imputation of missing data for income and earnings variables

For quantitative variables such as wages and total earnings, imputation was carried out rather than using special non-response codes. Imputation is the process by which missing or inconsistent items are "replaced" with plausible values. When carried out properly, imputation can improve data quality by reducing non-response bias. It also has the advantage of producing a complete data set for those variables being imputed.

The first step in the imputation process was a within-record imputation where missing information was replaced with values derived from the respondent's answer to other questions in the questionnaire using deterministic edit rules. In a few cases, "capping" was used, meaning that a respondent's answer was changed to a preset maximum or minimum allowable value for that variable. The remaining missing data were imputed using nearest-neighbor donor imputation. This is a widely used technique for treating item non-response. It aims at replacing missing information for a respondent with values provided from another respondent which is "similar" to him. Rules for identifying the respondent most similar to the non-respondent can vary depending on the variable being imputed. Donor imputation methods have good properties and generally will not alter the distribution of the data, which is a drawback of many other imputation techniques. Once the nearest neighbour imputation was done, within-record editing was performed again to ensure consistency of the data.

For the YITS, a list of the variables for which imputation was carried out can be found in the table that follows. Essentially, imputation was done for all of the earnings and income variables, for each of jobs 1 through 4 and job 7. The table shows an overall imputation rate for all jobs combined for each of the variables. Note that although imputation generally improves overall data quality, the artificial data created are used in estimation and can lead to underestimation of the sampling errors. This would only be a concern for variables with high imputation rates.

Table 7.6.1: Imputation rates for income and earnings

Variable	Number of records where questions apply	Number of imputed values	Imputation rate (%)
From section P:			
Hours worked per month	83,844	2,430	2.9
Weeks worked per month	83,844	666	0.8
Part-time/full-time flag	41,922	814	1.9
Earnings per hour	69,838	5,109	7.3
Earnings per month	69,838	5,110	7.3
Earnings per week	69,838	5,110	7.3
Earnings for self employed per year	6,112	1,264	20.7
From section U: Income from			
Employment insurance	1,506	379	25.2
Social assistance	1,004	126	12.5
Scholarships, grants and bursaries	3,326	61	1.8
Parents, gifts, etc. Not to be repaid	12,137	1,828	15.1
Other government sources	4,411	276	6.3
Non-government sources	1,318	221	16.8

Note: Earning variables were derived variables and were considered as being imputed if at least one of the components in deriving the earning was imputed.

7.7 Weighting

See Chapter 10 on Weighting.

8.0 Data Quality

8.1 The Frame

Given the important differences among the educational systems in Canada, the YITS 18-20 survey sought to generate estimates at the province level. For example, from the inception of the YITS, there was a clear desire for province-level estimates of high-school dropout rates among 20-year-old men and women and more generally, for estimates of proportions of at-risk groups by province. The study was also designed to support analysis of characteristics of at-risk groups at a regional level over possibly three cycles of a longitudinal panel. Interest in such characteristics influenced the data quality targets specified at the outset of work on the sample design. In concrete terms the objectives were:

- a CV less than or equal to 16.5% for Cycle 1 province-level leaver rates among 20-year-olds by sex
- at Cycle 3, for a characteristic found in a proportion of at least 20% of leavers, a CV less than or equal to 16.5% for that proportion at a regional level.

With some initial work on a sample design it became apparent these CV levels would be extremely difficult to attain for many of the province-level leaver rates. First, there was the question of a suitable frame to survey this relatively rare and highly mobile target population. Although this population comprises about 1.2 million individuals, they are found in only about one in ten households and have a one-year mobility rate of 22%.[11] With the proposed data quality levels it was estimated that if the sample were selected from an up-to-date household frame, the initial sample size would be about 20,000 persons. A sample selected from an older frame would have to be much larger to take into consideration the mobility rates of 18-20 year-olds.

Several options were examined in the search for a sample design based on a single frame. Unfortunately for the YITS, the frame used for the 1991 School Leavers Survey was no longer available because the Family Allowance program, a universally available benefit, had been replaced by the means-based Child Tax Credit. It should be noted that building a new frame was not an option for the YITS 18-20 cohort, due to the large amount of time and resources this would have required.

Single-frame designs based on the Labour Force Survey (LFS), the 1996 Census, random-digit dialling, electoral lists and tax files were considered, each with its own advantages and drawbacks. For the pilot survey a dual-frame dwelling-based design was proposed to combine the strengths of the LFS and Census frames. Central to the dual-frame sample design was the requirement to correctly classify respondents as movers or non-movers, according to whether their usual place of residence at the time of the YITS was different from that at the time of the 1996 Census. Analysis of the data from the pilot survey as well as mobility and school-leaving estimates from the 1996 Census data led us to reject the dual-frame sample design for the main survey. The primary factor in this decision was a projected large downward bias in important survey estimates such as leaver rates, resulting from a high incidence of mover misclassification.[12]

The final choice of frame for the 18-20 cohort main survey was the LFS. It was decided that a sample would be selected from persons in households currently in the LFS sample as well as those that had been in the LFS within the past three years. This restriction limited the YITS sample size, but it was important to recognise the potential risk of trying to trace the highly mobile target population in more dated LFS samples. The design required effective tracing and high response rates to control sampling error and potential bias in estimates for characteristics correlated with mobility. As for the data quality targets, various scenarios based on alternative assumptions for leaver rates, tracing success rates and response rates were examined. For example, leaver rates from the 1991 School Leavers Survey and the 1996 Census of Population

[11] Based on data from the 1996 Census of Population and the Labour Force Survey, 1997 to 1999.

[12] For more details on the sample designs considered see Moloney, Joanne C. and Hidiroglou, Mike C. (2000), Youth in Transition Survey: A Case Study in Design and Development, *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, pages 132-141.

were both included in these analyses. Although the data quality from the proposed design appeared unlikely to meet the initially proposed standards (as did all of the alternatives), the expected levels of sampling error and potential bias were judged to be acceptable.

8.2 Measuring sampling error

The estimates derived from this survey are based on a sample of households. The difference between the estimates obtained from the sample and the results from a complete count taken under similar conditions is called the sampling error of the estimate.

Since it is an unavoidable fact that estimates from a sample survey are subject to sampling error, sound statistical practice calls for researchers to provide users with some indication of the magnitude of this sampling error. This section of the documentation outlines the measures of sampling error which Statistics Canada commonly uses and which it urges users producing estimates from this microdata file to use also.

The basis for measuring the potential size of sampling errors is the standard error of the estimates derived from survey results.

However, because of the large variety of estimates that can be produced from a survey, the standard error of an estimate is usually expressed relative to the estimate to which it pertains. This resulting measure, known as the coefficient of variation (CV) of an estimate, is obtained by expressing the standard error of the estimate as a percentage of the estimate.

Much of the work on the sample design for the 18-20 cohort of the YITS focused on the desire to obtain reliable estimates of provincial leaver rates for 20-year-old men and women, as well as adequate numbers of respondents to support regional analysis of young adults in at-risk groups. The following tables relate to these survey objectives.

In general, the estimated CVs are consistent with the range predicted by scenarios examined under the final sample design. However, a quick look at Tables 8.2.1 to 8.2.4 indicates that analysis and interpretation of the YITS data should include statistical significance tests, and these must be based on estimates of sampling error that take into account the survey design. Caution is required particularly for analysis of rare characteristics in a population subgroup and for any characteristics within small subgroups.

Table 8.2.1. High-school leaver rates among 20-year-olds

Province	Men		Women	
	Estimate (%)	CV (%)	Estimate (%)	CV (%)
Newfoundland and Labrador	15.2	30.5	5.8	30.4
Prince-Edward Island	22.3	42.3	9.2	46.5
Nova-Scotia	14.5	21.5	5.0	37.1
New-Brunswick	11.7	24.8	3.7	38.7
Quebec	19.9	10.2	12.0	15.2
Ontario	11.2	13.3	7.8	13.6
Manitoba	15.7	20.6	13.9	18.2
Saskatchewan	9.9	17.6	4.5	27.7
Alberta	13.8	20.9	11.1	21.3
British-Columbia	16.7	21.6	8.9	20.2
All provinces	14.7	6.5	9.2	7.2

Table 8.2.2 Highest educational attainment of parents of high-school leavers

Region	Highest educational attainment of parents							
	Less than high school		High school		Post-secondary certificate/diploma		University degree	
	Estimate (%)	CV (%)	Estimate (%)	CV (%)	Estimate (%)	CV (%)	Estimate (%)	CV (%)
Atlantic	36.2	10.0	38.4	9.2	17.0	18.0	8.4	27.3
Quebec	29.9	10.0	44.2	6.8	19.6	13.0	6.3	21.4
Ontario	24.9	13.0	45.0	7.4	17.5	13.3	12.6	20.3
Prairies	26.4	11.9	47.7	6.3	12.7	15.9	13.2	17.4
British Columbia	18.6	19.6	48.1	11.3	15.0	21.7	18.3	21.8
All regions	26.9	5.9	45.2	3.6	16.9	7.2	11.0	9.7

Table 8.2.3 Self-assessment of high-school leavers' communication skills

Region	Skill level					
	Poor / fair		Good		Very good / excellent	
	Estimate (%)	CV (%)	Estimate (%)	CV (%)	Estimate (%)	CV (%)
Atlantic	25.2	10.4	40.3	8.2	34.6	9.6
Quebec	24.5	9.0	37.5	6.6	38.0	6.6
Ontario	17.6	13.0	40.0	7.1	42.4	7.3
Prairies	24.8	14.6	40.6	7.7	34.6	8.7
British Columbia	16.5	21.9	45.9	10.6	37.6	10.5
All regions	21.8	6.0	39.9	3.7	38.3	3.8

Table 8.2.4 Self-assessment of high-school leavers' problem-solving skills

Region	Skill level					
	Poor / fair		Good		Very good / excellent	
	Estimate (%)	CV (%)	Estimate (%)	CV (%)	Estimate (%)	CV (%)
Atlantic	21.4	10.5	52.6	6.2	26.0	11.1
Quebec	20.6	10.5	45.8	6.2	33.7	6.9
Ontario	14.8	13.8	52.9	5.5	32.3	8.7
Prairies	22.8	16.7	44.2	7.1	33.0	8.9
British Columbia	13.7	21.4	51.6	8.6	34.6	13.0
All regions	18.6	6.9	48.6	3.1	32.8	4.2

8.3 Non-sampling error

Errors that 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 and errors may be introduced in the processing and tabulation of the data. These are all examples of non-sampling errors. Considerable time and effort were used to reduce non-sampling errors in the survey. Quality assurance measures were implemented at each step of the data collection and processing cycle to monitor the quality of the data. These measures included the use of highly skilled interviewers, extensive training of interviewers with respect to the survey procedures and questionnaire, observation of interviewers to detect problems of questionnaire design or misunderstanding of instructions, procedures to ensure that data capture errors were minimised and coding and edit quality checks to verify the processing logic (see ***Chapter 6*** and ***Chapter 7***).

A major source of non-sampling errors in surveys is the effect of non-response on the survey results. The extent of non-response varies from partial non-response (failure to answer just one or some questions) to total non-response. Total non-response occurred because the interviewer was either unable to contact the respondent, or the respondent refused to participate in the survey. Total non-response was handled by adjusting the weight of households who responded to the survey to compensate for those who did not respond.

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

Partial non-response was low for the YITS. For example, when asked how many hours each week were spent on homework outside of class (F6), non-response was 0.3%. When asked a more sensitive question about drug use (FC30_F32), non-response was only 0.5%. Non-response for questions regarding grades obtained in school was as high as 3% (C0A, C8). As previously mentioned, partial non-response rates for income and earnings variables were higher and treated using nearest neighbour imputation. See section 7.6 on imputation for more detail on this process.

8.4 Response rates

Tables 8.4.1 to 8.4.3 present response rates for different subgroups of the sample. These rates are not weighted and use, as a base, the initial YITS sample count excluding a priori non-response households. In each of the 29,164 base households one person was selected and an attempt was made to contact and interview that person. The respondent count includes persons who were interviewed, persons contacted but confirmed to be outside the YITS target population by year of birth and persons whom a household contact confirmed as deceased. (The latter two groups are included in the respondent count because they provided all the relevant information, given their special status.)

The province-level response rates in Table 8.4.1 show considerable variation. There is also a definite pattern of higher response rates for sample units from more recent LFS groups (Table 8.4.3). Between the two demographic characteristics age and sex, age seems to be a more influential factor, although the age-sex response rates (Table 8.4.2) are more uniform than the province-level rates. All of these factors, among others, were examined for a possible role in the non-response adjustment of the weights (see Chapter 10 on weighting). Potential bias due to differing response rates by age and sex were also taken into account by the post-stratification step in the weighting process.

Table 8.4.1 Response rates by province of LFS household

Province	Response rate (%)
Newfoundland and Labrador	87.7
Prince-Edward Island	83.6
Nova-Scotia	83.4
New-Brunswick	79.7
Quebec	79.0
Ontario	78.9
Manitoba	84.5
Saskatchewan	84.2
Alberta	81.6
British-Columbia	80.4
All provinces	80.9

Table 8.4.2 Response rates by sex and age as of December 1999

Age	Response rate (%)		
	Men	Women	Both sexes
17-18	82.3	84.1	83.2
19	81.1	81.4	81.3
20-21	77.9	79.0	78.4
All ages	80.4	81.4	80.9

Table 8.4.3 Response rates by date of LFS household data

Date of LFS household data	Number of groups	Response rate (%)
1997 January-June	6	78.2
1997 July-December	5	78.5
1998 January-June	5	78.1
1998 July-December	6	80.0
1999 April-July	4	82.4
1999 August	4	85.4
1999 December	6	85.6
All dates	36	80.9

9.0 Guidelines for Tabulation, Analysis and Release

This section of the documentation outlines the guidelines to be applied by users tabulating, analysing, publishing or otherwise releasing any data derived from the survey microdata files. With the aid of these guidelines, users of microdata should be able to produce the same figures as those produced by Statistics Canada and, at the same time, will be able to develop currently unpublished figures in a manner consistent with these established guidelines.

9.1 Rounding Guidelines

In order that estimates for publication or other releases derived from these microdata files correspond to those produced by Statistics Canada, users are urged to adhere to the following guidelines regarding the rounding of such estimates:

a) Estimates in the main body of a statistical table are to be rounded to the nearest hundred units using the normal rounding technique. In normal rounding, if the first or only digit to be dropped is 0 to 4, the last digit to be retained is not changed. If the first or only digit to be dropped is 5 to 9, the last digit to be retained is raised by one. For example, in normal rounding to the nearest 100, if the last two digits are between 00 and 49, they are changed to 00 and the preceding digit (the hundreds digit) is left unchanged. If the last digits are between 50 and 99 they are changed to 00 and the preceding digit is incremented by 1.

b) Marginal sub-totals and totals in statistical tables are to be derived from their corresponding unrounded components and are then 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.

9.2 Sample Weighting Guidelines for Tabulation

The sample design used for the YITS was not self-weighting. When producing simple estimates including the production of ordinary statistical tables, users must apply the proper sampling weight.

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

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

9.2.1 Definitions of types of estimates: Categorical vs. Quantitative

Before discussing how the YITS data can be tabulated and analysed, it is useful to describe the two main types of point estimates of population characteristics that can be generated from the microdata file for the YITS.

Categorical Estimates

Categorical estimates are estimates of the number, or percentage of the surveyed population possessing certain characteristics or falling into some defined category. Whether the respondent has had any education in Canada or has worked more or less than 30 hours per week at a job are examples of such estimates. An estimate of the number of persons possessing a certain characteristic may also be referred to as an estimate of an aggregate.

Examples of Categorical Questions:

Q: Have you taken any elementary, junior high or high school education in Canada?
R: Yes / No

Q: Did you usually work 30 hours or more per week or less than 30 hours per week at this job?
R: 30 hours or more / Less than 30 hours

Quantitative Estimates

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

An example of a quantitative estimate is calculating the average number of hours worked per day by the surveyed population when they first started their job. The numerator could be an estimate of the total number of hours worked per week when they first started, and the denominator could be an estimate of the total number of days worked per week when they first started.

Examples of Quantitative Questions:

Q: When you first started this job, how many hours did you usually work per week?
R: |_|_| hours

Q: When you first started this job, how many days per week did you usually work?
R: |_|_| days

9.2.2 Tabulation of Categorical Estimates

Estimates of the number of people with a certain characteristic can be obtained from the microdata file by summing the final weights of all records possessing the characteristic(s) of interest. Proportions and ratios of the form X/Y are obtained by:

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

9.2.3 Tabulation of Quantitative Estimates

Estimates of quantities can be obtained from the microdata file by multiplying the value of the variable of interest by the final weight for each record, then summing this quantity over all records of interest. For example, to obtain an estimate of the total number of hours worked per week for those people who work part-time, multiply the value of total number of hours worked per week by the final weight for the record, and then sum this value over all records who reported working part-time.

To obtain a weighted average of the form X/Y, the numerator (X) is calculated as for a quantitative estimate and the denominator (Y) is calculated as for a categorical estimate. (Note: This applies when Y represents a subgroup of the survey population, but the characteristic Y could also be a quantitative estimate, as in the example above (in 9.2.1) for average number of hours worked in a day.) For example, to estimate the number of hours worked per week for those people who work part-time,
(a) estimate the total number of hours per week as described above,
(b) estimate the number of people in this category by summing the final weights of all records who reported working part-time
(c) divide estimate (a) by estimate (b).

9.3 Guidelines for Statistical Analysis

The YITS is based upon a complex sample design, with stratification, multiple stages of selection, and unequal probabilities of selection of respondents. Using data from such complex surveys presents problems to analysts because the survey design and the selection probabilities affect the estimation and variance calculation procedures that should be used. In order for survey estimates and analyses to be free from bias, the survey weights must be used.

While many analysis procedures found in statistical packages allow weights to be used, the meaning or definition of the weight in these procedures differ from that which is appropriate in a sample survey framework, with the result that while in many cases the estimates produced by the packages are correct, the variances that are calculated are poor.

For other analysis techniques (for example linear regression, logistic regression and analysis of variance), a method exists which can make the variances calculated by the standard packages

more meaningful, by incorporating the unequal probabilities of selection. The method re-scales the weights so that there is an average weight of 1.

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

- select all respondents from the file who reported SEX=male
- calculate the AVERAGE weight for these records by summing the original person weights from the microdata file for these records and then dividing by the number of respondents who reported SEX=male
- for each of these respondents, calculate a RESCALED weight equal to the original person weight divided by the AVERAGE weight
- perform the analysis for these respondents using the RESCALED weight.

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

For more information on calculating variance estimates for the YITS, see Chapter 11.

9.4 CV Release Guidelines

Before releasing and/or publishing any estimate from the YITS, users should first determine the quality level of the estimate. The quality levels are acceptable, marginal and unacceptable. Data quality is affected by both sampling and non-sampling errors as discussed in Chapter 8. However for this purpose, the quality level of an estimate will be determined only on the basis of sampling error as reflected by the coefficient of variation as shown in the table below. Nonetheless users should be sure to read section 8 to be more fully aware of the quality characteristics of these data.

First, the number of respondents who contribute to the calculation of the estimate should be determined. If this number is less than 30, the weighted estimate should be considered to be of unacceptable quality. (The figure 30 is for use with LFS based surveys and other surveys with generally small sampling fractions. From time to time, a lower figure may be appropriate for surveys with higher sampling fraction.)

For weighted estimates based on sample sizes of 30 or more, users should determine the coefficient of variation of the estimate and follow the guidelines below. These quality level guidelines should be applied to weighted rounded estimates.

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

Quality Level Guidelines

Quality Level of Estimate	**Guidelines**
1. Acceptable	Estimates have: a sample size of 30 or more, and coefficients of variation in the range 0.0% - 16.5% No warning is required.
2. Marginal	Estimates have: a sample size of 30 or more, and coefficients of variation in the range 16.6% - 33.3%. Estimates should be flagged with the letter M (or some similar identifier). They should be accompanied by a warning to caution subsequent users about the high levels of error, associated with the estimates.
3. Unacceptable	Estimates have: a sample size of less than 30, or coefficients of variation in excess of 33.3%. Statistics Canada recommends not to release estimates of unacceptable quality. However, if the user chooses to do so then estimates should be flagged with the letter U (or some similar identifier) and a warning should accompany the estimates.

10.0 Weighting

As the sample design of the YITS 18-20 cohort is determined to a large degree by the sample design of the LFS, so is the weighting.

10.1 Weighting procedures for the LFS

The final weight associated with each sample unit that contributes to the LFS estimates is the product of the following factors: the basic weight, the cluster sub-weight, the stabilisation weight, the non-response adjustment and a calibration adjustment to post-censal demographic estimates. These are described below.

10.1.1 LFS Basic weight

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

10.1.2 LFS Cluster sub-weight

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

10.1.3 LFS Stabilisation weight

Over time, application of the fixed cluster sampling rate to a growing population leads to an increasing sample size, as noted above. To control costs, dwellings are randomly dropped from the sample in order to maintain the desired sample size. This process, which is called sample stabilisation, changes a household's probability of being included in the sample. The stabilisation weight is the adjustment factor applied to the basic weight to reflect the change in the inclusion probability.

10.1.4 LFS Non-response adjustment

Notwithstanding the strict controls of the LFS, some non-response is inevitable, despite all the efforts made by the interviewers. For certain types of non-response (for example, household temporarily absent, refusal), data from a previous month's interview with the household (if these data exist), are brought forward and used as the current month's data for the household.

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

10.1.5 LFS sub-weight

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

10.1.6 LFS Calibration

The sub-weight can be used to derive a valid estimate of any characteristic for which information is collected by the LFS. In particular, estimates are produced for the total number of persons 15 years of age and older in provincial economic regions and the 24 large metropolitan areas as well as of designated age-sex groups in each of the ten provinces.

Independent estimates are available monthly for various age and sex groups by province. These are population projections based on the most recent Census data, records of births and deaths, and estimates of migration. In the final weighting step the sub-weights are adjusted so as to make corresponding survey estimates as consistent as possible with this auxiliary information. The technique used for this adjustment is known as regression estimation and it produces the final weights. Advantages of incorporating this weighting step include: consistency of estimates with demographic population estimates; an adjustment for coverage error; and reduction in the sampling error of estimates.

10.2 Weighting procedures for YITS

Because the sample for the YITS 18-20 cohort is derived from households that were in the LFS sample, part of the LFS weighting procedure is applicable to the YITS sample. More precisely, the LFS sub-weight is used as the initial weight for the YITS sample units. This weight represents the inverse of the probability of dwelling selection for the LFS, adjusted for LFS non-response.

To reflect the YITS sample design and take into account the non-response for this survey, several adjustments must be applied to the LFS sub-weight to derive the YITS final weight. In total, five adjustments are required to compensate for:

1. including 36 groups from the LFS
2. a priori non-response at the household level
3. selecting one person per household
4. non-response of the YITS selected persons
5. differences between post-censal demographic estimates and the weighted counts derived from the first four adjustments

10.2.1 YITS Adjustment for 36 groups

As noted in the section on LFS weighting, the sample comprising each of the six LFS rotation groups can be treated as a representative sample of the population. In fact, many surveys that use the LFS sample do select fewer than six rotation groups, say *r* groups. In such cases, the adjustment factor *6/r* is normally applied to the LFS sub-weight to compensate for the reduced probability of selection. The analogous adjustment in the case of the YITS would be *6/36*.

However, this simple adjustment assumes the sample in each of the groups is statistically independent, which is not true in the case of the YITS. A sample dependency among the 36 YITS groups arises from the appearance of a given cluster in two or more LFS groups that rotate out six (or a multiple of six) months apart. The households in these groups represent different random starts (that is, sub-samples) within the same cluster, and as such, they are not independent samples. To the extent that clusters are homogeneous, there could be a strong positive correlation among estimates of characteristics based on these groups of households. For such characteristics, ignoring the dependency among groups in the weighting process would lead to an underestimation of the sampling error.

The weighting method implemented for the 18-20 cohort of the YITS includes an adjustment to take account of the rotation patterns of individual LFS clusters. In essence, the adjustment assigns less weight to sample units that arise from different groups but belong to the same cluster. To create a set of weights independent of the characteristics to be analysed, the procedure assumes a fixed correlation of 1. Some characteristics of interest for the YITS may not have a high positive correlation among estimates from the same cluster – for such characteristics, the estimates of sampling error based on these weights will be somewhat conservative, that is, overstated.

10.2.2 YITS Household non-response

To account for a priori non-response households in the initial YITS sample, the step 1 weights (that is, weights after the adjustment for 36 groups) of the remaining households were proportionally increased so that the sum of their adjusted weights would equal the sum of the step 1 weights of all households in the initial YITS sample. The adjustment was done separately within response classes defined by the following variables:

- province of LFS household
- month and year of LFS household data (grouped into 7 time periods)
- type of LFS stratum: urban (including the apartment frame) or rural (including the remote areas)

This adjustment is based on the assumption that households included on the YITS sample file for collection represent the characteristics of all households in the initial YITS sample.

10.2.3 YITS Sampling within household

The final stage of sampling for the YITS was the selection of one person among household members in the target population. To reflect the probability of selection within the household, the step 2 weight, that is, the weight after adjustment for household non-response, was multiplied by the number of household members in the target population.

10.2.4 YITS Non-response of selected person

Among the 29,164 selected persons for whom an interview was sought, 23,592 (80.9%) were classified as respondents.13 To account for person-level non-response the step 3 weights of the respondents were proportionally increased so that the sum of their adjusted weights would equal the sum of the step 3 weights of all selected persons. Although the overall non-response rate was quite low, it was apparent from analysis at the sample design stage that the non-response adjustment should take into consideration the varying vintage of the LFS contact information among the 36 YITS groups and, if possible, the mobility of the sample units. The adjustment was done separately within response classes defined by the following variables:

- province of the LFS dwelling
- month and year of LFS household data (grouped into 7 time periods)
- age of the selected person as of December 31, 1999
- traced/mover status: whether the selected person had to be contacted at a different phone number than the one on the LFS frame

This adjustment is based on the assumption that respondents represent the characteristics of all selected persons in the YITS sample.

10.2.5 YITS Post-stratification

The final weighting adjustment for the YITS 18-20 cohort is post-stratification. The purpose of this adjustment is similar to that of the calibration adjustment for the LFS weights.

The adjustment was applied within post-stratification classes defined by year of birth, sex and province. Official post-censal demographic projections for the LFS target population are created for single-year age groups, but not by year of birth. The control totals used for the YITS are averages of population counts derived from the LFS age control totals for the 12 months of 1999.

Among the 23,592 respondents treated in the non-response adjustment, the post-stratification adjustment was applied to the subset of 22,378 who were born in 1979 to 1981 and who agreed to share their data with the client department. The step 4 weights of these respondents were adjusted proportionally so that the sum of their final weights would be equal to the control total in their respective post-stratification classes.

[13] This count includes cases that were contacted and confirmed to be deceased or outside the YITS target population by year of birth

11.0 Variance estimation

Due to the complexity of the YITS sample design a re-sampling technique was chosen to calculate estimates of the variance. For the 18-20 cohort, the bootstrap re-sampling method is used. This technique is popular among surveys with a large number of strata and multiple primary sampling units (PSU) per stratum. Unlike the Jackknife method the bootstrap does not suffer from inconsistent estimates for population estimates such as percentiles.

The idea behind the bootstrap method is to select random sub-samples from the full sample in such a way that each of the sub-samples (or replicates) follows the same design as the full sample. The final weights for units in each replicate are recalculated, following the same weighting steps used for the full sample (see ***Chapter 10***). These bootstrap weights are used to calculate a population estimate for each replicate. The variance among the replicate estimates for a given characteristic is an estimate of the sampling variance of the full-sample population estimate.

For the YITS 18-20 cohort a total of 1,000 replicates were created. These are called BSW1 to BSW1000 on the data file. Each replicate was formed by sampling independently within each stratum. If there were *n* PSUs in a stratum, (n-1) were selected by simple random sampling with replacement. While sampling with replacement to create the bootstrap samples is a departure from the full-sample design for the YITS, this is a common practice in large surveys with small first-stage sampling fractions because it greatly simplifies the variance estimation process at the expense of overestimating the true variance slightly.[14]

[14] As noted in the weighting steps, the adjustment for common clusters among the 36 groups may also contribute to an overestimation of the sampling variance for characteristics that do not have a high positive correlation within the cluster