



**Youth in Transition Survey (YITS)
Cohort A - 17-year-olds
Cycle 2**

USER GUIDE

Revised December 2006



Statistics
Canada Statistique
Canada

Canada

TABLE OF CONTENTS

1.0 Introduction.....	4
2.0 Background.....	5
2.1 YITS Component.....	5
2.2 Reference Documentation.....	6
2.3 Objectives.....	6
3.0 Concepts and Definitions.....	7
3.1 YITS Concepts.....	7
3.1.1 Move to U.S./Return to Canada (Module A).....	8
3.1.2 Education and School Activities – high school, junior high or elementary (Modules B, C, D and F).....	8
3.1.3 Postsecondary Education (Modules H, K, M and N).....	12
3.1.4 Postsecondary Engagement (Module K).....	15
3.1.5 Financing Postsecondary Education (Modules L and M):.....	15
3.1.6 Education and Work Aspirations (Module M).....	15
3.1.7 Health – Activity Limitations (Module M).....	15
3.1.8 Support from Others (Module N).....	15
3.1.9 Employment (Modules P1, P2 and P5).....	16
3.1.10 Courses or Training Programs Related to Job or Career (Module P6).....	21
3.1.11 GAPS - Months not in school full-time and not working (Module PS).....	21
3.1.12 Volunteering (Module Q).....	21
3.1.13 Skills (Module R).....	22
3.1.14 Personal Characteristics and Family Background (Module U).....	22
3.1.15 Income (Module V).....	24
4.0 Data Collection and Processing.....	26
4.1 The Questionnaire.....	26
4.2 Training.....	26
4.3 Supervision and Control.....	26
4.4 The Interview.....	26
4.5 Follow-up of Non-response.....	27
4.6 Data capture.....	27
4.7 Minimum completion requirements.....	27
4.8 Computer Assisted Interview (CAI) Editing.....	27
4.9 Head Office Processing.....	28
5.0 Derived Variables and Codebooks.....	31
5.1 Cycle 2 Codebooks.....	31
5.1.1 Person Level Main File Codebook.....	32
5.1.2 Education Above High School – Institution Roster Codebook.....	43
5.1.3 Education Above High School – Program Roster Codebook.....	44
5.1.4 Postsecondary Engagement Roster Codebook.....	46
5.1.5 Job Roster Codebook.....	47
5.1.6 Job Details Roster Codebook.....	48
6.0 YITS Scales.....	49
6.1 Defining Scales and Their Uses.....	49
6.1.1 What is a Scale?.....	49
6.1.2 Why Use a Scale?.....	50
6.1.3 What Type of Scales are Used in YITS?.....	50
6.1.4 Response Biases.....	51
6.1.5 Negative and Positive Questions.....	51
6.2 Scale Development.....	51
6.2.1 Investigation of Model Validity.....	52

6.2.2	Estimation of the Scores.....	53
6.2.3	Scale Reliability	54
6.2.4	Testing Scale Validity	55
6.3	The High School Engagement Scale	56
6.3.1	Why is there Interest in a School Engagement Model?	56
6.3.2	What is the YITS High School Engagement Scale?	56
6.3.3	Structural Model of High School Engagement	57
6.3.4	Description of High School Engagement.....	57
6.4	Analysis of High School Engagement Scale	59
6.4.1	Testing Model Validity	60
6.4.2	Estimating Scores.....	63
6.4.3	Scale Reliability	67
6.4.4	Testing Scale Validity	67
6.5	Social Support Scale	68
6.5.1	Description of Social Support Scale	68
6.5.2	Model Validity	69
6.5.3	Estimating Scores.....	69
6.5.4	Scale Reliability	70
6.5.5	Testing Scale Validity	70
6.6	Scale References	71
7.0	Survey Methodology	73
8.0	Data Quality.....	74
8.1	The Frame	74
8.2	Measuring Sampling Error	74
8.3	Non-sampling Error.....	76
8.4	Response rates	78
9.0	Imputation of Missing Data for Income and Earnings Variables.....	79
10.0	Guidelines for Tabulation, Analysis and Release	81
10.1	Rounding Guidelines	81
10.2	Sample Weighting Guidelines for Tabulation.....	82
10.2.1	Definitions of types of estimates: Categorical vs. Quantitative	82
10.2.2	Tabulation of Categorical Estimates	83
10.2.3	Tabulation of Quantitative Estimates	83
10.3	Guidelines for Statistical Analysis	83
10.4	CV Release Guidelines	84
11.0	Weighting.....	86
11.1	Cycle 2 Student Weight.....	86
11.2	Cycle 2 Parent Weight	86
11.3	Sub-Domain Weights	86
12.0	Variance Estimation	88
13.0	Working with YITS files.....	89
13.1	Roster and Flat Files.....	89
13.2	Youth In Transition Survey: Data Extraction Tool.....	89
13.2.1	About Youth in Transition Survey (YITS).....	89
13.2.2	Statistical Activity.....	90
13.2.3	Purpose of the Application.....	91
13.2.4	Saving and Loading Queries.....	97
APPENDICES.....		98
LINKS TO REFERENCE DOCUMENTS.....		101

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 second data collection – Cycle 2 of the survey - for the cohorts aged 17 and 20 to 22, was conducted by Statistics Canada between February and May 2002 with the co-operation and support of Human Resources and Skills Development Canada (HRSDC).

This User Guide for cycle 2 of the YITS 17 year-olds is developed for the first release of the microdata file. Throughout this document, this cohort may be referred to as Cohort A or Cohort 1, 15-year-old Reading Cohort or 17 year-olds (their age as of December 2001).

A separate User Guide has been written for the Cycle 2, Cohort B - 18-20 year-olds who were 20-22 years-old as of December 2001.

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
150 Tunney's Pasture Driveway
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 and Skills Development Canada:

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

2.0 Background

Starting in 1996, Human Resources, Skills Development Canada (HRSDC), 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 parents¹. 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 years-old at the time of data collection. The Canadian YITS was first administered alongside the Programme for International Student Assessment (PISA) in 2000. This cohort is referred to as the “YITS 15 year-old Reading Cohort”.

For further information on PISA, an international dataset, which includes Canadian data and full documentation for this dataset, can be found under www.pisa.oecd.org.

2.1 YITS Component

YITS is designed to examine the patterns of, and influences on, major transitions in young people's lives, particularly with respect to education, training and work. Human Resources and Skills Development Canada and Statistics Canada have been developing the YITS in consultation with provincial and territorial ministries and departments of labour and education. Content includes all formal educational experiences and most labour market experiences, achievement, aspirations and expectations.

The results from the Youth in Transition Survey have many uses. Human Resources and Skills Development Canada can use them to aid policy and program development. Other users of the results include educators, social and policy analysts, and advocacy groups. The information will show how young adults are making their critical transitions into their adult years.

These researchers and analysts will have access to important information that can be used in developing programs to deal with both short-term and long-term problems or barriers that young adults may face in their pursuit of higher education or in gaining work experience. Information from the survey will help to evaluate the effectiveness of existing programs and practices, to determine the most appropriate age at which to introduce programs, and to better target programs to those most in need.

Young adults themselves will be able to see the impact of decisions relating to education or work experiences. They will be able to see how their own experiences compare to those of other young adults.

¹For 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 and Skills Development Canada, 2000, Cat. No. MP32-30/00-5E/F)

2.2 Reference Documentation

Reference documentation for both cohorts for Cycle 1 and Cycle 2, can be found on the Statistics Canada website www.statcan.ca. The Meta database provides information for both cohorts under 'Definitions, Data Sources and Methods':

- The 15 year-old Reading Cohort (2000) PISA (Survey 5060) and YITS (Survey 5058);
- 18-20 year-old cohort YITS (Survey 4435)

2.3 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 Postsecondary 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 postsecondary entry and postsecondary 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".

Objectives for Cycle 2 and future cycles are:

- to confirm data at attendance of educational institutions collected in previous cycle; and
- to follow youth as they move to accommodate the attendance at educational institutions and acceptance of employment

3.0 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 2001. 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 3.1 describes the concepts included in the YITS and Section 5.0 describes derived variables included on the data file and in codebooks.

3.1 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.

Cycle 2 of YITS brings forward historical data from Cycle 1 and asks respondents to confirm the information or to update to previous information (names or dates). Comparison can then be made between the cycles' different reference periods (Cycle 1 – as of December 1999 and Cycle 2 – January 2000 to December 2001).

Sub-Section	Description	Modules
3.1.1	Move to the United States/Return to Canada	A
3.1.2	High school, junior high or elementary education Status Secondary School courses Secondary School activities School – attitudes and behaviours	B C D F
3.1.3	Postsecondary education	H, K and M
3.1.4	Postsecondary School Engagement	K
3.1.5	Financing Postsecondary Education	L and M
3.1.6	Loans and education and work aspirations	M
3.1.7	Health – activity limitations	M
3.1.8	Support from others	N
3.1.9	Employment: Labour market and job roster Employment details Reservation Wage	P1 P2 P5
3.1.10	Courses or Training Programs Related to Job or Career	P6
3.1.11	GAPS - months not in school full-time and not working	PS
3.1.12	Volunteering	Q
3.1.13	Skills	R
3.1.14	Personal Characteristics and Family Background	U
3.1.15	Income	V

3.1.1 Move to the U.S./Return to Canada (Module A)

Respondents who live in the United States, or lived there anytime between January 2000 and December 2001 were interviewed. They may have moved to attend school, to work or to accompany a parent or spouse/partner. Visits and temporary vacation periods were not included. The dates of moves are essential to assist in clarifying factors such as education funding, level of income, educational levels based on institutions and programs enrolled in and support systems. Respondents may have had more than one move to and from the U.S., for example if they attended school and then returned to work in Canada during breaks. As well, they may have had periods of work in both the United States and Canada. Information on resident status is also collected.

Temporary Resident Status: the respondent may be living in the U.S. for a specified period to attend school on a student visa, to be in training with a business, or to work for a company (also referred to as “non-immigrant” status).

Permanent Resident: the respondent is entitled to remain in the U.S. for an undetermined period and to work and/or attend school (also referred to as a ‘green card holder’).

3.1.2 Education and School Activities – high school, junior high or elementary (Modules B, C, D and F)

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 mathematics.

Respondents are first asked to confirm information collected for the Cycle 1 reference period (as of December 1999), then to provide information on their current enrolment status (January 2000 to December 2001). 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: 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 : 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: 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: respondents who were continuing their studies at elementary, junior high or high school in December 2001. Respondents continuing with postsecondary are not classified as high school continuers.

Correspondence courses (Distance Education): 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: 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 2000 and 2001.

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
Nunavut	5 yr-old kindergarten	1-6	7-9	10-12

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

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

Graduates – high school: 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: 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 postsecondary non-university education.

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

Province	Graduation diploma, certificate or equivalent
Newfoundland and Labrador	<ul style="list-style-type: none"> →Senior High Graduation Diploma →Provincial High School Graduation Certificate →High School Equivalency Diploma →General Educational Development (GED)
Prince Edward Island	<ul style="list-style-type: none"> →High School Graduation Certificate →General Educational Development (GED)
New Brunswick (English and French Sectors)	<ul style="list-style-type: none"> →High School Diploma or Certificate →Adult High School Diploma →General Educational Development (GED)
Nova Scotia	<ul style="list-style-type: none"> →High School Graduation Diploma →N.S. Secondary School Equivalency Certificate →General Educational Development (GED)
Quebec	<ul style="list-style-type: none"> →Diplôme d'études secondaires (DES)/Secondary School Diploma (SSD, general education) →Diplôme d'études professionnelles (DEP)/Secondary School Vocational Diploma (SSVD) →Attestation de formation professionnelle (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	<ul style="list-style-type: none"> →Ontario Secondary School Diploma (OSSD) →General Educational Development (GED)

Province	Graduation diploma, certificate or equivalent
Manitoba	<ul style="list-style-type: none"> →Provincial Diploma →Adult Basic Education Certificate →General Educational Development (GED)
Saskatchewan	<ul style="list-style-type: none"> →Record of Secondary Level standing →High School Equivalency Diploma →General Educational Development (GED)
Alberta	<ul style="list-style-type: none"> →Alberta High School Diploma →Certificate of Achievement →High School Equivalency Diploma for Mature Students →General Educational Development (GED)
British Columbia	<ul style="list-style-type: none"> →Certificate of Graduation or “Dogwood Diploma” →School Completion Certificate →General Educational Development (GED)
Northwest Territory	<ul style="list-style-type: none"> →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)
Nunavut	At this time Nunavut was still using the Alberta graduation standards
Yukon	<ul style="list-style-type: none"> →Yukon Graduation Certificate →School Completion Certificate

Junior high / Intermediate / Middle school: 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: respondents who had not completed the high school graduation requirements, and were not attending elementary, junior high or high school in December 2001.

Private Elementary/ Secondary (high) school: 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): This is a category for the highest level of education completed in Module B for respondents taking schooling in Québec. It is considered to be at the high school level. In cycle 1, only a few respondents reported this in Module B, whereas the information was reported and captured in Module H as to the type of program a respondent was working towards, or in which they 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 programs: programs or courses, combined with high school that 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.

3.1.3 Postsecondary Education (Modules H, K, M and N)

In Module H, respondents are first asked to confirm information on education from Cycle 1, to determine eligible institutions and programs for the reference period 2000 to 2001 which they may:

- still be enrolled in;
- have graduated from;
- have left; or
- have changed (program or main field of study).

Eligible institutions and programs from Cycle 1 begin the roster of postsecondary education in Module H and are referred to as "open".

The postsecondary education sections contain questions on the respondent's participation in any education, above the high-school level and more than 3 months in duration, that could be counted towards a degree, certificate or diploma from an educational institution. An **eligible postsecondary 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 2002.

All respondents, who in December 2001, were no longer in high school, or who had completed the high school graduation requirements were asked if they had taken any postsecondary education prior to January 2002.

The order of 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 derived variable such as HLATTD2 that indicates the status of the respondent at that institution as being a "continuer" or "non-continuer" (see section 5.0, Derived Variables).

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

Module H collects information on the type of postsecondary 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.

Youth, who report being enrolled in a postsecondary program, are asked 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

using Classification of institutions and programs (CIP)² coding. CIP codes are available on the data file. The youth also report on attitudes and behaviours while taking postsecondary education.

Apprenticeship Programs: 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.

College Post-Diploma or Graduate Program: 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” or “a certificate in international business administration”.

Commercial school: 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: 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 – Postsecondary: respondents who were continuing their studies towards the completion of a postsecondary program in December 2001.

Degrees - first professional: 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 a first professional degree, rather it is a professional licence.

Degrees versus Diplomas: 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: 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: to be deemed eligible, a postsecondary 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 2002. 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.

² See section 4.10

Ever dropped out: question is asked of:

- Postsecondary graduates who at some point dropped out of their program, but returned to continue their education until graduation;
- Postsecondary continuers who at some point dropped-out of their program but returned to continue their education; and
- school leavers - those who never graduated.

Fellowship: A Fellowship is a position in a university held by a graduate student having teaching duties as part of his or her educational program.

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

Graduates – Postsecondary: respondents who have completed the graduation requirements towards a diploma, certificate or degree.

Leavers / Non-completers – Postsecondary: 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 2001.

Licence (Québec): Licence, Licentiate, Testamur are credentials awarded mainly by religious programs in Quebec. The term 'Licence' does not include professional association licences, and are to be specified under "other". Interviewers were asked to identify whether the licence was at a graduate or post-graduate level.

On-the-job experience program (Module H): programs or courses, combined with postsecondary 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 that are profit oriented and are engaged in providing professional and vocational training, and are licensed by the province.

Programs – Postsecondary (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.

Trade /vocational certificate or diploma: 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 upholstery. Many community colleges or technical institutes offer certificates or diplomas at the trade level.

University: an independent institution granting degrees in at least arts and sciences.

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

3.1.4 Postsecondary Engagement (Module K)

For Cycle 2, Module K was revised to accommodate the possible collection of zero, one or two sets of postsecondary engagement questions. The goal, originally, was to acquire information with respect to the respondent's first postsecondary experience. For many of the respondents, this would simply be the first institution above the high school level that they have attended. However, based on information already obtained during the YITS Cycle 1 data collection, it was found that students attending CEGEP institutions and students attending NON-CEGEP institutions (e.g., university, community college outside of Québec, etc) have two distinct postsecondary experiences (in terms of cost and distance away from home). This motivated the decision to attempt to collect two sets of postsecondary engagement questions for students studying in Québec (1 CEGEP and 1 first non-CEGEP).

3.1.5 Financing Postsecondary Education (Modules L and M):

All respondents who had taken a postsecondary program were asked about their sources of income and the amounts of income used to finance their postsecondary education.

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

Grants: 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.

Scholarships, award or prizes : 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.

Funding also includes government-sponsored student loans, money from family or relatives that does not have to be paid back; money from trust funds, RESPs or RRSPs; and/or money from jobs or from personal savings.

3.1.6 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.

3.1.7 Health – Activity Limitations (Module M)

All respondents provide information about any long term physical condition(s), mental condition or health problem(s) 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.

3.1.8 Support from Others (Module N)

Support from others has an impact on success in education and in the labour market. This section asks that respondents consider types of support they may receive.

3.1.9 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 of this cohort are still in full-time schooling or have left not too long ago, measuring work activities during the current cycle's reference period represented a good alternative. For those who had left full-time schooling prior to 2001, additional questions on the first job after leaving full-time schooling were added in order to capture the "transition" job. In addition, for those who had not worked in the current cycle reference period, 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, in addition to the first job after leaving full-time schooling, if this event occurred prior to 2001. "Open" jobs from the previous cycle were confirmed for eligibility. Start and stop dates were collected for a maximum of six jobs the respondent worked at between January 2000 and December 2001. (Module P1). This job count included previous cycle "open" and eligible jobs and the current cycle eligible jobs.

Open jobs were those at which a respondent worked during the previous cycle. They may have still been employed but not actually working at the job prior to January 2000 (temporary layoff, business slowdown, etc.). Respondents are asked if they are still working with that employer in Cycle 2 and/or whether they had returned to work for the employer. (See Eligibility, job).

Further job details were collected on a maximum of four jobs the respondent worked at during the reference period (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 2001 Pilot with youth aged 20-22) indicated that approximately 2% of youth held more than four jobs over a one-year period.

Apprenticeship: The employer undertakes, by contract, to employ and train an apprentice under the supervision of a qualified journeyman. To become an apprentice, there is usually a formal registration process with a provincial Ministry or Department or Trade organization.

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 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.

Class of Worker: There are three main categories of worker that are defined further on in this section – Paid Worker (an employee); Self-employed Worker; Unpaid worker in the family farm or business.

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 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 if 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 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 2002. 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 2001 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 2000 to 2001 reference period must meet the following criteria. If the respondent was still working at that job from Cycle 1 (“open” job) in the Cycle 2 reference period or if the job began in 2000 or 2001, 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 31, 2001. Any eligible jobs would remain “open” for the next cycle.

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 31, 2001 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 as of December 31, 2001 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 as of December 31, 2001 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.

Job/Work (Modules P1, P2, P5, P6 and PS) Any activity carried out by the respondent during the reference period for pay or profit, includes 'payment in kind' (payment in goods or services rather than money) whether actual payment was received during the reference period. Work includes time spent:

- actively looking for work, clients, preparing estimates, quotes or tenders, establishing a new business;
- operating a business, professional practice or farm even if no sales were made, no professional services were rendered or nothing was actually produced;
- as the owner or manager of a business even though the person is physically unable to do the actual work;
- on-the-job training; unpaid work for a family business or farm; odd jobs.

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 2001 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 2001. (See **Job losers** and **Job leavers**.)

Reservation Wage (Module P5): The questions asked in this section are required to understand if a respondent is willing to work just for money, or wants to have work that will pay them what they feel they are worth (from other job experience or from acquired educational skills). They may feel that although there are better job opportunities in another locale, they would not be willing to move to improve their job or career prospects.

Self-employed (Modules P1 and 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 do not have a business but who are paid directly by a client such as a child care giver; house cleaner; dog walker - with one or more clients, who provides these services on a formal or informal contractual basis.
- 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 (Modules P1 and 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. Excluded are respondents who perform regular household chores around the house or yard (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 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**).

Workfare (Ontario and Destination Emploi (Quebec)): are provincial programs in which participants exchange their labour services for social assistance payments. Such a worker is classified as an “employee”.

3.1.10 Courses or Training Programs Related to Job or Career (Module P6)

Regardless of whether a respondent worked in the last two years, they are asked for information on any courses or training programs related to a job or career. These programs might be sponsored by an employer or taken to have better job opportunities in a current job, or in the labour market. If a program was made ineligible in Module H, the respondent is asked in Module P6 if it might be applicable. A maximum of four courses or training programs are flagged for collection (2 for employer organized/sponsored and 2 for any other training related to a job or career).

Training, outside of formal educational programs and training courses taken to acquire skills for a job or career, might include reading books, manuals or other written materials, using materials available electronically; or watching others work, receiving advice or assistance from others.

3.1.11 GAPS - Months not in school full-time and not working (Module PS)

For the reference period of January 2000 to December 2001, 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 2000 and 2001, when the respondent was not in school full-time (high school or postsecondary).

For every gap month, the respondent was asked what he/she was 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(s) of activities.

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

3.1.12 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.

3.1.13 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.

3.1.14 Personal Characteristics and Family Background (Module U)

Citizenship : 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 naturalisation (citizenship process)

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 : refers to the ethnic or cultural group(s) to which the respondent's ancestors belong.

Dependent children: 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. This includes birth, adopted, step and foster children from the previous and current YITS cycle.

Household: 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: 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: spoken well enough to conduct a conversation - languages in which the respondent can carry on a conversation at some length on various topics.

Language: spoken at home and home and most often – persons who live alone were asked to report the language that they feel most comfortable speaking, others it would be mostly their 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: marital status (conjugal status) of respondent at time of interview. Marital status from cycle 1 is confirmed if respondent indicated either married or living common-law. For cycle 2 the respondent is asked, if there is a change in their marital status, how many relationships they have entered into since January 2000 and the date when their marital status changed. The categories are as follows:

Single

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

Married

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

Living Common-law or with a partner (girlfriend, boyfriend)

Persons who are not legally married but are sharing the same usual address and living as husband and wife.

Widowed

Persons who have lost their spouse through death 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.

Parents or guardians: The respondent is asked about the relationship to their parents or guardians and living arrangements during high school, as well as their interest in the respondent's future education or career options

Permanently moved out: Respondents who have moved away from their family home and have a permanent address different from that residence are considered to have permanently moved out. They may, in time move back, or move to different addresses.

Province: the data file contains the province that the respondent considers to be their address (variable NDPR_D2); the province in which the respondent last took high school (Module B), and the province where they attended a postsecondary institution (Module H). The province variable is also located in the Entry Component

Spouse or partner background: If a respondent is legally married or living common-law, they are asked about their spouse's/partner's highest level of education completed and for their spouse's/partner's current main activity. Details of kind of business are also required if the respondent's spouse or partner is employed.

Usual Place of Residence:

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: 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.

3.1.15 Income (Module V)

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 and Skills Development Canada (employment insurance program).

Farm self-employment NET income: This is receipts minus operating expenses, depreciation and capital costs allowances. If it is a partnership, the respondent only reports their own share. If the farm is incorporated, the income is reported in Wages and Salaries and/or Dividends. Net rent from farms/property leased to others is reported in Other Investment Income.

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: 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); and Worker's Compensation.

Income from other non-government sources: 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: 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: income from scholarships, grants or bursaries may be provided either to the respondent or to the institution to pay for tuition. 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: 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: 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: 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. Earnings received from all paid worker jobs held during the year are to be reported.

4.0 Data Collection and Processing

Data collection for Cycle 2 of the YITS took place between February and May 2002.

4.1 The Questionnaire

The YITS questionnaire was developed for cycle 1 using CASES software for a computer assisted telephone interview (CATI). In cycle 2 the software was changed from CASES to BLAISE. 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.

4.2 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 second phase of the training took place in the regional offices. Senior interviewers and project managers, who were trained during the first phase, provided the same training to interviewers in the regional offices.

4.3 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.

4.4 The Interview

In Cycle 1 a single respondent was 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 58 minutes. The total average questionnaire time (Modules A through Z) was 46 minutes.

4.5 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 two types of non-response:

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.

Partial interviews

If a respondent failed to complete the interview, the case was assigned a partial code. Callbacks to convert the case to a complete were only made if the survey response rate was not met or a particular segment of the target population experienced low response rates (as determined by Methodology). For example, Cohort B, high school-leavers .

4.6 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.

4.7 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 Derived Variables and Codebooks - **Section 5.0**).

4.8 Computer Assisted Interview (CAI) Editing

CAI editing occurred in the Regional Offices during data collection. The data were collected in a telephone interview using a CAI application. As such, it was possible to build various edits and checks into the questionnaire in order to ensure high quality of the information collected. Below are specific examples of the types of edits used in the YITS computer-assisted interviewing application:

Review Screens (Confirmation Screens)

Review screens were created for important and complex information. For example in Module H both an Institution Confirmation Screen and Program Confirmation Screen were used to identify eligible institutions and programs. In Module P1, the Job Confirmation Screens identified eligible jobs. All review screens provide essential information to assist interviewers and respondents in recall and verification.

Range Edits

Range edits were built into the CAI system for questions asking for numeric values. If values entered were outside the range, the system generated a pop-up window that stated the error and instructed the interviewer to make corrections to the appropriate question. Range edits were provided for years, number of months, weeks, days or hours of work, and monetary values through out the questionnaire. For example, please see B_Q20Y. This question asked the respondent what year they were last in high school and the acceptable range was 1983 – 2001. If the respondent answered outside of this range the interviewer was to prompt for a more accurate date.

Flow Pattern Edits

All flow patterns were automatically built into the CAI system. For example, in Module A, at A_Q01, respondents were asked if they had lived in the United States at any time in the last two years. If not, they flowed to Module B for questions on high school status. Whereas the group that lived in the U.S. were asked a number of questions related to their stay in the U.S. before going to Module B.

Consistency Edits

Consistency edits (indicated within a variable name as “E” or “D”) were used to identify inconsistencies in responses. These edits also identified missing information from previous variables. They were included as part of the CAI system to allow interviewers to return to previous questions to correct for inconsistencies. Interviewer instructions were displayed for handling or correcting problems such as incomplete or incorrect data. Variable B_E46e asks the respondent if they have completed the SSVD graduation requirements. This question is asked to correct SSVD graduation status.

4.9 Head Office Processing

The main outputs of the Youth in Transition Survey (YITS), Cycle 2 are "clean" data files. This section presents a brief summary of some of the processing steps involved in producing these files.

The processing of the YITS Cycle 2 data was done using the Generalized Processing Environment. This is a generic system that follows a series of steps to “clean” a file from beginning to end. The main steps were:

- Clean-up
- Pre-edit
- Flow edits
- Coding
- Derived variables
- Computer Generated Edits

To facilitate processing, the file was split into smaller files, which corresponded to the survey modules. Rostered files were also created in order to process the data more efficiently.

The YITS prepared both flat files and rostered files for Cycle 2. The rostered data files contain as many records for a given longitudinal respondent as the number of events, such as for the number of employers and the employment history, the number of institutions attended and programs taken at those institutions. (The employer questions would be programmed in a roster, which would be, repeated the appropriate number of times.)

Clean Up

The purpose of this step is to drop full-duplicate records and split-off records with duplicate identification numbers for examination. The data is then split between response and non-response based on pre-determined criteria. For more information, see *Section 7.0*.

Pre-edits

For all records where values were missing (blank) from the collection, the value of “9, 99, 999...” was inserted during Head Office processing to indicate that no information was collected. The “Don’t know” values returned by the CAI application as code “9” are changed to “7” in the pre-edits. As well, the “Mark all that apply” questions were de-strung and values converted to Yes (1) or No (2) responses. Finally, all text answers were removed from the processing file and set aside to be handled separately.

Recoding Other Specify

A number of questions in each module of the YITS have a category of “other specify”. Interviewers manually type in a “long answer” response that was not easily categorized during the interview or could not be coded into one of the already pre-assigned categories for that question. During the pre-edit, the “other specify” responses were reviewed and when possible coded back up to already existing categories, or new categories were created, when they met a specific criteria. For example, in B_Q47

and B_Q52 (for which the derived variable DRED2 was created) there was a high frequency of "moved" as a response under the other category.

Questions containing a response category of ""other specify" may not be consistent across cycles. While the original response categories as appeared on the questionnaire are always included it is possible that additional categories may be generated based on the frequency of responses found in the "other specify" category.

General rules were applied:

1. If a new category existed in a previous cycle, it will retain the same category value in the current cycle.
2. A new category in a cycle will always retain its value regardless of whether it is being used in a future cycle. The "other-specify" (general) would take on the last value. Users should be cautious when comparing "other specify" counts across cycles as they may not always contain similar items.

For example:

For variable X we have values 1-5 and "other" is 6. After Other Specify Long Answer (OSLA) coding two new categories were added and are given the values 6 and 7 and "other" moves down to 8. Next cycle, if we only use new category 6, and 7 is left off, the "other" stays as 8. In this context the value "8" is dissimilar across cycles.

Flow Edits

Standards have been developed for the coding structure of data in order to explain certain situations in a consistent fashion. These standard codes are applied at the flows editing stage of processing of the YITS. The following describes these various situations and the codes used to describe the situation.

Valid Skip

In some cases a question, a series of questions, or an entire module was not applicable to the survey respondent. For example, a respondent reporting he/she has no children will have all questions related to dependent children set to a valid skip during processing. A code "6", "96" "996" was used on the data file to indicate that a question is a valid skip. In cases where an entire module of the questionnaire was not applicable to the respondent, all the variables of the module were set to a "Valid Skip".

Not stated

The not stated code indicates that the answer to the question is unknown. Not stated codes were assigned for the following reasons.

- As part of the CAI interview, the interviewer was permitted to enter a "Refusal" or "Don't know" code. When this happened, the CAI system was often programmed to skip out of this particular section of the questionnaire. As part of the Generalized Processing System, all of the subsequent questions of this section are assigned a "Not stated" code. A not stated code means that the question was not asked of the respondent. In some cases it is not even known if the question was applicable to the respondent.
- In some cases a questionnaire was started but ended prematurely. For example, there may have been some kind of an interruption, or the respondent decided that she/he wished to terminate the interview. If a questionnaire was only partially completed but enough information had been collected to consider to meet the preset criteria to stay in the sample, then all remaining unanswered questions on the questionnaire were set to "Not stated". The one exception was that if it was known that a certain question or section was not applicable, then these questions were set to "Valid Skip".

An item which was coded as "Not stated" is indicated by a code "9". For a variable that is one digit long the code is "9", for a two-digit variable "99", for a three-digit variable "999", etc.

Coding of Open-ended Questions

A few data items on the YITS questionnaire were recorded by interviewers in an open-ended format. For example, in the Employment Section (Module P1), respondents who had worked in the reference period were asked a series of open-ended questions regarding their employment:

- What kind of business, service or industry is/was this?
- What kind of work are/were you doing?
- At this work, what are/were your most important duties or activities?

The answer provided by the respondent was entered as text by the interviewer. At Head Office, the open-ended questions were coded using various standard classifications. Occupation questions were coded using the 1991 Standard Occupational Classification codes (SOC) and the industry questions were coded using the 1997 North American Industry Classification System (NAICS). Other open-ended questions for education (type of institution and program) were also collected and codes were assigned for Education (CIP).

Imputation

For a few variables on the YITS file, rather than using a special non-response code, imputation was carried out. Imputation is the process whereby missing or inconsistent items are "filled in" with plausible values. For the YITS, imputation was carried out for sections of the survey pertaining to employment income and household income. See *Section 8.0* for more details on imputation.

Derived Variables

Once the data are "clean", derived variables (see *Section 5.0*) are programmed following specifications written by subject matter staff of (CTCES) and programmed by the processing team of Special Surveys Division. The standard codes of valid skip and not stated are also applied in the Derived Variable Stage. In a few instances, a category has been set to 5, 95, 995 or 9995. This reserve code refers to Not Applicable. Furthermore, if one or more of the input variables (to the derived variable) had a "Refusal", "Don't know" or "Not stated" code, then the derived variable was set to "Not stated".

Final Processing Files

The final processing files created for the YITS, Cycle 2 included both flat and rostered files. Users should see *Section 13.0* for working with the file.

5.0 Derived Variables and Codebooks

Answers from certain questions from the survey were combined to form specific concepts. Among them are: high school student status; postsecondary school status; and labour force status in December 2001. Such concepts are called *derived* variables (DVs). 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 following tables identify the derived variables and the respective codebooks. The derived variables are listed under specific headings and in the order in which they appear in the codebooks. There are modules for which derived variables were not required. The universe for each derived variable indicates who responded to the questions contained in the derived variables, and may be for both cohort A (15 year-olds – referred to as 17 year-olds in Cycle 2) and cohort B (18-20 year-olds – referred to as 20-22 year-olds in Cycle 2), or either cohort. In the cases where a DV is for cohort B only, it will not be listed in this User Guide but can be found in the Cycle 2, 20-22 year-old User Guide.

5.1 Cycle 2 Codebooks

The codebooks for Cycle 2 were developed depending on whether a module contained historical data (brought forward from Cycle 1) or were Cycle 2 reference period only. For those modules containing Cycle 1 and Cycle 2 data, rostered codebooks were produced. The following is a table of codebooks all of which can be found on the Statistics Canada website at www.statcan.ca under the Youth in Transition Survey (project 4435), cycle 2.

Section	Codebook Name	File Type	Descriptive File Name	Modules	Cycle
5.1.1	Main_cohortb_eng.doc	Flat file	Person Level Main File	A, B, C, D, F, L, M, N, Q, R, U, V	Cycle 2
5.1.2	Hinst_cohortb_eng.doc	Institution Roster	Education above High School	H	Cycle1 and 2
5.1.3	Hprog_cohortb_eng.doc	Program Roster	Education above High School	H	Cycle 1 and 2
5.1.4	Modulek_cohortb_eng.doc	Roster	Post Secondary Engagement Roster	K and KE	Cycle 2
5.1.5	P1cycle1_eng.doc	Roster	Confirmation of Open Jobs from Cycle 1 (cohort b only)	P1	Cycle 1
5.1.6	P2main_cohortb.doc	Roster	Job Roster	P1 and P2	Cycle 2
5.1.7	P1cycle2_eng.doc	Roster	Job Details Roster	P1	Cycle 2

5.1.1 Person Level Main File Codebook

Module Name and identifier	Derived variable	Description	Universe
Entry	RecordID	Respondent identification number	All respondents
	AGED2	Age of respondent as of Dec. 2001	(as above)
	BYEARD2	Respondent's year of birth (1979 – 1984)	(as above)
	BMONTHD2	Respondent's month of birth	(as above)
	GENDERD2	respondent's gender	(as above)
	PROVD2	Province Code of the province of residence for the household as of date of interview (2001 Census geography)	(as above)
	URMIZD2	An indicator of the rural or urban geography, based on the MIZ coding of the census subdivision. A MIZ is Metropolitan Influence Zone - based on the percentage of people who commute to an urban core	(as above)
Moved to the United States – Module A	No derived variables		
High School Status – Module B	ACMD2	Respondent's age at date completed high school diploma requirements or equivalent	Respondents who have completed high school diploma or equivalent.
	AFTD2	Respondent's age when last in elementary or secondary school full-time prior to January 2002	All respondents
	AHSD2	Respondent's age when last in elementary or secondary school prior to January 2002	(as above)
	DNOD2	Number of times respondent dropped out of elementary or secondary school	(as above)
	DRED2	Main reason for having ever dropped out of elementary or secondary school prior to January 2002	Respondents who have ever dropped out of school regardless of whether they were a high school continuer, graduate or leaver as of December 2001. For respondents who have dropped out more than once, this variable is the main reason for the last time they have dropped out
	DRPD2	Variable indicating if respondent has ever dropped out of elementary or secondary school	All respondents
	FPTLESD2	Full-time/part-time status when respondent was last in elementary or secondary school	(as above)
	HG9D2	Variable to indicate if respondent has gone past High School Status	(as above)
	HGCD2	Highest grade respondent has completed in elementary or secondary school as of December 2001	(as above)
	HSSTATD2	High school status as of December 2001	All respondents except those who in cycle 1 did not have any high school, junior high or elementary education in Canada or the United State and who in cycle 2, did not take any high school, junior high or elementary education

Module Name and identifier	Derived variable	Description	Universe
	LESMTD2	Date (month) respondent was last in elementary or secondary school prior to January 2002	All respondents
	LESYRD2	Date (year) respondent was last in elementary or secondary school prior to January 2002	(as above)
	LFTESMD2	Date (month) respondent was last in elementary or secondary school full-time prior to January 2002	(as above)
	LFTESYD2	Date (year) respondent was last in elementary or secondary school full-time prior to January 2002	(as above)
	LGED2	Grade enrolled in elementary or secondary school as of December 2001 or the date last in school (elementary or secondary)	Respondents who have not completed the minimum requirements for a high school graduation certificate, diploma or its equivalent as of December 2001 or the date last in school. Excludes those last enrolled in an SSVD program in Quebec
	HSDIPMD2	Date (month) respondent completed high school diploma requirements or equivalent	Respondent who completed their high school diploma, Secondary School Vocational diploma (SSVD) requirements or equivalent
	HSDIPYD2	Date (year) respondent completed high school diploma requirements or equivalent	(as above)
	NUMHSD2	Number of different high schools attended between January 2000 and December 2001	All respondents
	SCIPD2	Respondent's main field of study or specialization - secondary school vocational diploma program	Respondents who took some education in an SSVD program between January 2000 and December 2001
	SCIPRD2	Respondent's field of study or specialization - secondary school vocational diploma program (primary grouping)	(as above)
Secondary School Courses – Module C	HGCLD2	Highest grade completed - main language course between January 2000 and December 2001	Respondents who have taken any elementary, junior high or high school education between January 2000 and December 2001
	HGCMD2	Highest grade completed - mathematics course between January 2000 and December 2001	Respondents who have taken any elementary, junior high or high school education between January 2000 and December 2001
	LLLD2	Level of last main language course - priority destination between January 2000 and December 2001	(as above)
	LLMD2	Level of last mathematics course - priority destination between January 2000 and December 2001	(as above)
	PLLD2	Province completed last high school (junior	(as above)

Module Name and identifier	Derived variable	Description	Universe
		high) main language course between January 2000 and December 2001	
	PLMD2	Province completed last high school (junior high) mathematics course between January 2000 and December 2001	(as above)
Secondary School Activity – Module D	THEWEPD2	Total number of hours spent with an employer during a work experience program	Respondents who were in high school between January 2000 and December 2001 and who participated in a work experience program
High School Engagement – Module F	HPDHSD2	Average number of hours of paid in a week during last year of high school	Respondents who took high school, junior high or elementary school courses between September 2000 and December 2001, INCLUDING those who never attended an educational institution (i.e. had done their studies through home schooling or distance education)
	HUWHSD2	Average number of hours of unpaid work in family's business or farm during last year of high school	Respondents who took high school, junior high or elementary school courses between September 2000 and December 2001, INCLUDING those who never attended an educational institution (i.e. had done their studies through home schooling or distance education), and who between January 2000 and December 2001 lived in a household where someone in the family owned a family business or farm
High School Engagement – Scales and Standard Errors – Module F	YSHACPS2	This variable measures a respondent's academic participation at high school	All 17 year-old respondents who took high school, junior high or elementary school courses between September 2000 and December 2001, excluding those who had never attended an educational institution (i.e. had done their studies through home schooling or distance education)
	YSHACPE2	Standard error for the score on high school academic participation	(as above)
	YSHAC1S2	This variable measures a respondent's academic identification with high school	(as above)
	YSHAC1E2	Standard error for the score on academic identification with high school	(as above)
	YSHSOES2	This variable measures a respondent's social engagement at high school	(as above)
	YSHSOEE2	Standard error for the score on social engagement at high school	(as above)

Module Name and identifier	Derived variable	Description	Universe
	YSHACES2	This variable measures a respondent's academic engagement at high school	(as above)
	YSHACEE2	Standard error for the score on academic engagement at high school	(as above)
	YSHSCES2	This variable measures a respondent's overall engagement at high school	(as above)
	YSHSCEE2	Standard error for the score on overall engagement at high school	(as above)
Education and Training – Module H	AGSPSD2	Respondent's age at start of first postsecondary program	All respondents except high school continuers who had not graduated
	DLFPSMD2	Date (month) respondent was last taking postsecondary education on a full-time basis ever, prior to January 2002	(as above)
	DLFPSYD2	Date (year) respondent was last taking postsecondary education on a full-time basis ever, prior to January 2002	(as above)
	DLPSMD2	Date (month) respondent was last taking postsecondary education ever, prior to January 2002.	Respondents who took some postsecondary education between January 2000 and December 2001
	DLPSYD2	Date (year) respondent was last taking postsecondary education ever, prior to January 2002.	(as above)
	EDTPSMD2	Date (month) respondent first started postsecondary education prior to January 2002	(as above)
	EDTPSYD2	Date (year) respondent first started postsecondary education prior to January 2002	(as above)
	FPSPD2	Variable identifying respondent's first postsecondary institution and program attended	All respondents except high school continuers who had not graduated
	HEDATD2	Highest certificate, diploma or degree attained (or graduated from) as of December 2001	(as above)
	HEDLD2	Highest education level taken as of December 2001	(as above)
	HGDAD2	Highest graduation diploma attained as of December 2001	All respondents
	HLPSD2	Highest level of postsecondary education taken across all programs and institutions as of December 2001	All respondents except high school continuers who had not graduated
	LPSATD2	Overall postsecondary status as of December 2001	(as above)
	MHSPSD2	Duration of time, in months, from the date last in elementary/secondary school to the time started first postsecondary program. MHSPSD2 is duration in absolute value. MHSPSFD2 indicates whether duration is negative or positive	Respondents who took some postsecondary education
	MHSPSFD2	Duration of time, in months, from the date last in elementary/secondary school to the time started first postsecondary program. MHSPSFD2 indicates whether duration is negative or positive. MHSPSD2 is duration	(as above)

Module Name and identifier	Derived variable	Description	Universe
		in absolute value	
	NEPRCD2	Number of eligible postsecondary programs taken between January 2000 and December 2001	Respondents who attended at least one postsecondary program between January 2000 and December 2001
	NINDID2	Number of postsecondary institutions attended between January 2000 and December 2001	(as above)
Funding – Module L	TMOD2	Total amount of money owed to student loans, bank loans, lines of credit, parents or family as of December 2001, to fund postsecondary education	Respondents who received a loan from the government, a bank, a line of credit or from parents or family as of December 2001 in order to fund their postsecondary education
	TMRD2	Total amount of money received from scholarships, awards, prizes, grants or bursaries between January 2000 and December 2001, to fund postsecondary education	Respondents who received a scholarship, award, prize, grant or bursary between January 2000 and December 2001 in order to fund their postsecondary education
Loans and Aspirations – Module M	JOA30RD2	2 digit occupation code (SOC 1991) for job or occupation respondent would be interested in having around the age of 30	All respondents
	JOA30D2	4 digit occupation code (SOC 1991) for job or occupation respondent would be interested in having around the age of 30	(as above)
	M1BMD2	Date (month) respondent was last in school full-time	(as above)
	M1BYD2	Date (year) respondent was last in school full-time	(as above)
Social Support Scales and Standard Error – Module N	YSHSUPS2	This variable measures how much social support a respondent receives from friends, family and other sources	All 17 year-old respondents
	YSHSUPE2	Standard error for the score on how much social support respondent receives from friends, family and other sources	(as above)
Work related Questions – Module P1	ELJBVD2	Indicates the number of eligible jobs (during this cycle) up to a maximum of 7 jobs (based on first 6 jobs + first job after leaving full-time schooling)	All respondents
	ELJBND2	Contains the remaining number of jobs worked at (during this cycle) for which no verification was possible	(as above)
	FJ_AGED2	Respondent's age when started the first job s/he worked at after leaving full-time schooling	Respondents who were in full-time school between January 2000 and November 2001; respondents aged 17 who were last in school full-time prior to January 2002; respondents aged 20-22 who were last in school full-time prior to January 2002 who had not yet had a first

Module Name and identifier	Derived variable	Description	Universe
			job after leaving full-time school (as above)
	FSTJOB2	Identifies the job the respondent held at time of leaving full-time schooling, or the job first started after leaving full-time schooling (jobs 1 to 7 collected), for those respondents who were no longer full-time students in December 2001	(as above)
	FTSFJDD2	Duration of time, in months, from the date left full-time schooling to the time started first job after full-time schooling. FTSFJDD2 is duration in absolute value. (Use in combination with FTSJFLD2 to determine if duration is positive or negative)	Respondents who began their first job after leaving full-time schooling between January 2000 and December 2001
	FTSJFLD2	Determine if the value of FTSFJDD2 is positive or negative	(as above)
	JBST01D2 to JBST24D2	Employment status - Flag for each month (January through December) in 2000 and 2001 indicating if respondent was employed at at least one job	Respondents who had at least one job between January 2000 and December 2001
Reservation Wage – Module P5	RSWGD2	Reservation wage - the lowest wage or salary a respondent would accept to begin a new job immediately, full-time - expressed in dollars and cents per hour	All respondents
Training – Module P6	ERTD2	Number of employer related training courses or programs taken between January 2000 and December 2001	Respondents who worked during the reference period (January 2000 to December 2001)
	CRTD2	Number of career or job related training courses or programs taken between January 2000 and December 2001	All respondents
	TTH_1D2 and TTH_2D2	Total number of training hours - Employer related training - courses 1 and 2	Respondents who reported at least one training course or program taken between January 2000 and December 2001
	TTH_3D2 and TTH_4D2	Total number of training hours- Job/career related training - courses 1 and 2	(as above)
	TTHERD2	Total number of training hours- Employer related training	(as above)
	TTHJCD2	Total number of training hours- Job/career related training	(as above)
	TTHD2	Total number of training hours	(as above)
	CT1CID2 and CT2CIPD2	Respondent's main field of study or specialization - career-related training - courses 1 and 2	Respondents who reported at least one career-related training course taken in the last two years (between January 2000 and December 2001)
	ET1CIPD2 and ET2CIPD2	Respondent's main field of study or specialization - training organized by an employer - courses 1 and 2	Respondents who reported at least one training course or program organized by an employer taken in the last two years (between January 2000 and December 2001)
	C1CIPRD2 and	Respondent's main field of study or specialization - career-related training	Respondents who reported at least one career-related training

Module Name and identifier	Derived variable	Description	Universe
	C2CIPRD2	(primary grouping) - courses 1 and 2	course taken in the last two years (between January 2000 and December 2001)
	E1CIPRD2 and E2CIPRD2	Respondent's main field of study or specialization - training organized by an employer (primary grouping)	Respondents who reported at least one training course or program organized by an employer taken in the last two years (between January 2000 and December 2001)
GAPS – Module PS	FEDS01D2 to FEDS24D2	Full-time student status - Flag for each month in 2000-2001 indicating if respondent was a full-time student	All respondents
	FTES01D2 to FTES24D2	Whether the respondent was in elementary, secondary or postsecondary full-time schooling for each month during 2000-2001	Respondents who took some education (elementary, secondary school or postsecondary education) between January 2000 and December 2001
	LGMD2	Last Gap Month. Last month in the reference period when the respondent was not employed at a job and was not a student full-time	Respondents who had at least one month in which they were not working or at school between January 2000 and December 2001
	LWLGMD2	Looking for Work Last Gap Month. Variable to indicate if respondent was looking for a job during their last gap month of the reference period	Respondents who had at least one month in which they were not working or at school between January 2000 and December 2001
	MLTJBYD2	Flag to indicate if respondent was employed at 2 or more jobs during one month or more at any time between January 2000 and December 2001	Respondents who had at least one job between January 2000 and December 2001
	MTJOB01D2 to MTJOB24D2	Variable identifies, for each month from January 2000 to December 2001, the number of jobs the respondent was employed at during the month	Respondents who had at least one job between January 2000 and December 2001
	REAS01D2 to REAS24D2	Respondent's main activity during month when not employed at a job - created for each month in 2000-2001 where respondent was not employed in at least one job	Respondents who had at least one month in which they were not working or at school between January 2000 and December 2001
	WSTP01D2 to WSTP24D2	Working at job status - Flag for each month in 2000-2001 indicating if respondent was employed and working at at least one job	Respondents who had at least one job between January 2000 and December 2001
	WKFPTPD2	Full-time/part-time status for respondent at all jobs in December 2001	(as above)
Volunteer Activities – Module Q	OVRD2	Respondent's volunteer activities in 2001	All respondents
Skills – Module R	No derived Variables		
Demographics –	VISMND2	The concept of visible minority applies to	All respondents

Module Name and identifier	Derived variable	Description	Universe
Module U		persons 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	
	CITZEND2	This variable updates the information on citizenship collected in the previous cycle. It is possible that respondents may have obtained Canadian citizenship or acquired or relinquished citizenship from other countries	All respondents
	LANIMMD2	Whether or not respondents who are not Canadian by birth have ever been a landed immigrant	Respondents who reported that they were not Canadian by birth
	YRMCD2	Year respondent first came to Canada to live	(as above)
	ATMD2	Respondent's approximate age at time of moving to Canada	(as above)
	FAMSTRD2	Parent(s) or guardian(s) the respondent lived with MOST of the time during high school (or junior high or elementary school if respondent has not done high school)	All respondents. For respondents aged 20-22 (cohort 2), the data have been brought forward from cycle 1
	PGATHD2	Parent or guardian in the family home, when the respondent was living at home or away, during most of high school	All respondents
	PG1D2 to PG4D2	parent/guardian – first to fourth (if required)	All respondents
	AGMPD2	Respondent's age, in years, at time of moving out permanently from the home of parents or guardians	Respondents who reported that they have moved out permanently from the home of their parents or guardians as of December 2001
	MARSTAD2	Marital status	All respondents
	SPINDD2	4 digit industry code (NAICS 1997) for spouse's job	Respondents who reported a partner/spouse. Information was collected for some respondents who reported a status of common-law in Cycle 1 and stated that they were still living with the same person in Cycle 2 but reported their marital status as single
	SPINDRD2	2 digit industry code (NAICS 1997) for spouse's job	(as above)
	SPOCCD2	4 digit occupation code (SOC 1991) for spouse's job	(as above)
	SPOCCRD2	2 digit occupation code (SOC 1991) for spouse's job	(as above)
	DEPCHD2	Number of dependent children	All respondents
	CBDYM1D2 to CBDYM4D2	Date (month) of birth of all dependent children	All respondents who reported dependent children

Module Name and identifier	Derived variable	Description	Universe
	CBDYY1D2 to CBDYY4D2	Date (year) of birth of all dependent children (child 1 to 4)	(as above)
	RELCH1D2 to RELCH4D2	Relationship of dependent children to respondent (child 1 to 4)	Respondents with dependent children
	LVECH1D2 to LVECH4D2	Status of living arrangement of dependent children in the household (one to four)	(as above)
	HPMCH1D2 to HPMCH4D2	Reason dependent children live with respondent most or part of the time (child 1 to 4)	Respondents with dependent children who live in the same house with the child most or part of the time
	LCCCD2	Language(s) spoken well enough to conduct a conversation	All respondents
	LSMHD2	Language(s) spoken most often at home	(as above)
Income - Module V	INCEID2	Income received in 2001 from Employment Insurance benefits	(as above)
	INCEISD2	Income respondent's spouse or partner received in 2001 from Employment Insurance benefits	Respondents who had a partner or spouse in 2001
	INCGSD2	Income received in 2001 from other government sources such as Worker's Compensation or Canada Pension Plan or Quebec Pension Plan	All respondents
	INCGSSD2	Income respondent's spouse or partner received in 2001 from other government sources such as Worker's Compensation or Canada Pension Plan or Quebec Pension Plan	Respondents who had a partner or spouse in 2001
	INCNRD2	Income received in 2001 from parents or other people that did not have to be repaid (excludes loans).	All respondents
	INCNRSD2	Income respondent's spouse or partner received in 2001 from parents or other people that did not have to be repaid (excludes loans)	Respondents who had a partner or spouse in 2001
	INCOND2	Income received in 2001 from other non-government sources including income from dividends, interest and other investment income, employer pensions, annuities or rental income	All respondents
	INCONSD2	Income respondent's spouse or partner received in 2001 from other non-government sources including income from dividends and other investment income, employer pensions, annuities or rental income	Respondents who had a partner or spouse in 2001
	INCSCD2	Income received in 2001 from spousal support or child support	All respondents
	INCSCSD2	Income respondent's spouse or partner received in 2001 from spousal support or child support	Respondents who had a partner or spouse in 2001
	INCSED2	Income received in 2001 from self-employment	All respondents
	INCSESD2	Income respondent's spouse or partner received in 2001 from self-employment	Respondents who had a partner or spouse in 2001
	INCSGD2	Income received in 2001 from scholarships, grants or bursaries	All respondents

Module Name and identifier	Derived variable	Description	Universe
	INCSGSD2	Income respondent's spouse or partner received in 2001 from scholarships, grants or bursaries	Respondents who had a partner or spouse in 2001
	INCSPD2	Income received in 2001 from Social Assistance or Provincial Income Supplements	All respondents
	INCSPSD2	Income respondent's spouse or partner received in 2001 from Social Assistance or Provincial Income Supplements	Respondents who had a partner or spouse in 2001
	INCSTD2	Income received in 2001 from the Goods and Services Tax Credit (GST) or Harmonized Sales Tax Credit (HST) or Quebec Sales Tax Credit (QST)	All respondents
	INCSTSD2	Income respondent's spouse or partner received in 2001 from the Goods and Services Tax Credit (GST) or Harmonized Sales Tax Credit (HST) or Quebec Sales Tax Credit (QST)	Respondents who had a partner or spouse in 2001
	INCTBD2	Income received in 2001 from Canada Child Tax Benefit or provincial child tax benefits or credits	All respondents
	INCTBSD2	Income respondent's spouse or partner received in 2001 from Canada Child Tax Benefit or provincial child tax benefits or credits	Respondents who had a partner or spouse in 2001
	INCWSD2	Income received in 2001 from wages or salaries	All respondents
	INCWSSD2	Income that respondent's spouse or partner received in 2001 from wages or salaries	Respondents who had a partner or spouse in 2001
	TINCD2	Total 2001 income from all sources before taxes and deductions	All respondents
	TINCSD2	Total 2001 income for respondent's spouse or partner from all sources before taxes and deductions	Respondents who had a partner or spouse in 2001
	DECEASE2	Respondents deceased in 2000 or 2001	
	Weight	Weight : decimal in 4th byte of the field	

The following geography derived variables were not released in the codebook

CMA25D2- Area consisting of one or more adjacent municipalities situated around a major urban core. To form a census metropolitan area, the urban core must have a population of at least 100,000. To form a census agglomeration, the urban core must have a population of at least 10,000.

CMACAD2 - Identifies the type of census metropolitan area (CMA) or census agglomeration (CA) in which the enumeration area is located. This field will be left blank where the enumeration area is not part of a CMA or CA.

CSDCD2 - Census subdivision (CSD) is the general term for municipalities (as determined by provincial legislation) or areas treated as municipal equivalents for statistical purposes (for example, Indian reserves, Indian settlements and unorganized territories).

EAD2 - Each enumeration area is assigned a three-digit code that is unique within a federal electoral district (FED). In order to identify each EA uniquely in Canada, the two-digit province/territory code and the three-digit FED code must precede the EA code.

ECRD2 - Economic Region (ER) of residence for the household at time of interview

SATYPD2 - Category assigned to a municipality not included in either a census metropolitan area (CMA) or a census agglomeration (CA). (A CMA or CA is an area consisting of one or more adjacent municipalities situated around a major urban core. To form a CMA, the urban core must have a

population of at least 100,000. To form a CA, the urban core must have a population of at least 10,000.) A municipality is assigned to one of four categories depending on the percentage of its residents who commute to work in the urban core of any census metropolitan area or census agglomeration. Strong MIZ: more than 30% of the municipality's residents commute to work in any CMA or CA. Moderate MIZ: from 5% to 30% of the municipality's resident's commute to work in any CMA or CA. Weak MIZ: from 0% to 5% of the municipality's residents commute to work in any CMA or CA. No MIZ: fewer than 40 or none of the municipality's residents commute to work in any CMA or CA

REGD2 – Region of residence for the household as of date of interview

5.1.2 Education Above High School – Institution Roster Codebook

Module Name and identifier	Derived variable	Description	Universe
Entry	RecordID	Respondent identification number	All respondents
	INST_ID	This number given to the institution corresponds to the order in which the respondent reported it	Respondents with at least one postsecondary institution
Education and Training – Module H	DSAINMD2	Date (month) respondent started postsecondary education at this institution, prior to January 2002	Respondents who took some postsecondary education between January 2000 and December 2001
	DSAINYD2	Date (year) respondent started postsecondary education at this institution, prior to January 2002	(as above)
	DLINMD2	Date (month) respondent was last at this institution between January 2000 and December 2001	(as above)
	DLINYD2	Date (year) respondent was last at this institution between January 2000 and December 2001	(as above)
	HLATTD2	postsecondary status at this institution as of December 2001	(as above)
	NEPRPID2	Number of eligible postsecondary programs taken at this institution between January 2000 and December 2001	(as above)
	DLFINMD2	Date (month) respondent was last taking postsecondary education at this institution on a full-time basis prior to January 2002	(as above)
	DLFINYD2	Date (year) respondent was last taking postsecondary education at this institution on a full-time basis prior to January 2002	(as above)
	FPLIND2	Full-time or part-time student when last at this institution between January 2000 and December 2001	(as above)
	TYPEID2	Type of postsecondary institution	(as above)
	PSCMD2	Campus Code	(as above)
	PSPROVD2	Province of postsecondary institution	(as above)
	INSCDD2	Institution Code	(as above)
	PSIPOSID	Postsecondary institution position identifier which identifies the cycle and position where the data in this cycle for this program was collected	Respondents with at least one postsecondary institution
	PSILNGID	Postsecondary institution longitudinal identifier which permits following an institution across cycles	Respondents with at least one postsecondary program at this institution

The postsecondary institution longitudinal identifier, PSILNGID, allows the user to follow an institution across the cycles. The identifier has 2 digits which can take the following values:
 First digit: represents the cycle in which the institution information was first collected; can take values of 1 or 2
 Second digit: represents the institution attended by respondent; can take values of 1 to 4
 This identifier does not exist on the cycle 1 data file. If you need to retrieve a variable from the cycle 1 file you will need to refer to the table on Appendix B which indicates which variables belong to which institution. For example if you need question H8, and PSILNGID=12 then variable H8b, is associated with institution 2.

5.1.3 Education Above High School – Program Roster Codebook

Module Name and identifier	Derived variable	Description	Universe
Entry	RecordID	Respondent identification number	All respondents
	INST_ID	This number given to the institution corresponds to the order in which the respondent reported it	Respondents with at least one postsecondary institution
	PROG_ID	This number given to the program within the institution corresponds to the order in which the respondent reported it	(as above)
Education and Training – Module H	INELGHD2	Flag indicating whether or not an open program and/or institution from cycle 1 was deemed ineligible in cycle 2 education at this institution, prior to January 2002	Respondents with an open program/institution from cycle 1
	LVPRD2	Level of postsecondary program	Respondents who took some postsecondary education between January 2000 and December 2001
	CLGPRD2	Postsecondary status in this program as of December 2001	(as above)
	DSPRMD2	Date (month) respondent started this postsecondary program, prior to January 2002	(as above)
	DSPRYD2	Date (year) respondent started this postsecondary program, prior to January 2002	(as above)
	DLPRMD2	Date (month) respondent was last taking this postsecondary program between January 2000 and December 2001	(as above)
	DLPRYD2	Date (year) respondent was last taking this postsecondary program between January 2000 and December 2001	(as above)
	FPLPRD2	Full-time or part-time student when last in this program, between January 2000 and December 2001	Respondents who were taking a postsecondary program between January 2000 and December 2001
	DLFPRMD2	Date (month) respondent was last taking this postsecondary program, on a full-time basis prior to January 2002	Respondents who took some postsecondary education between January 2000 and December 2001
	DLFPRYD2	Date (year) respondent was last taking this postsecondary program, on a full-time basis prior to January 2002	(as above)
	SIPRD2	For postsecondary programs which are ongoing or completed as of December 2001, whether respondent has stopped or interrupted their education between January 2000 and December 2001	(as above)
	AGEPSD2	Respondent's age at start of postsecondary program	Respondents who were taking a postsecondary program between January 2000 and December 2001
	NUMDURD2	Time spent taking a postsecondary program, as of December 2001 (months)	Respondents who participated in a postsecondary program between January 2000 and December 2001
	RSIPRD2	For programs in which respondents participated between January 2000 and December 2001, reason for stopping or interrupting program if the respondent stopped or interrupted their program position identifier which identifies the cycle and position where the data in this cycle for this program	Respondents who took some postsecondary education in programs which are ongoing or completed who have ever stopped or interrupted their studies

Module Name and identifier	Derived variable	Description	Universe
		was collected	
	CIP1D2	Respondent's first main field of study or specialization	Respondents who had a valid postsecondary program
	CIP2D2	Respondent's second main field of study or specialization	(as above)
	CIP1RD2	Respondent's first main field of study or specialization (primary grouping)	(as above)
	CIP2RD2	Respondent's second main field of study or specialization (primary grouping)	(as above)
	THEPSD2	Total time spent with an employer in a co-op, apprenticeship, trade/vocational training or another program (e.g. practicum, internship or clinical) for this program	Respondents who attended an eligible postsecondary program between January 2000 and December 2001 and participated in a program which included on the job experience and/or time spent in a workplace
	OPSPD2	Chronological order of postsecondary programs attended by respondent during 2000 and 2001	Respondents who took some postsecondary education between January 2000 and December 2001
	PSPPOSID	Postsecondary program position identifier which identifies the cycle and position where the data in this cycle for this program was collected	Respondents with at least one postsecondary program
	PSPLNGID	Postsecondary program longitudinal identifier which permits following a program across cycles	(as above)
	ICYID	Postsecondary institution cycle identifier, which identifies the cycle where data were first collected for this institution	(as above)

The postsecondary program longitudinal identifier, PSPLNGID, allows the user to follow a program across the cycles. The identifier has 4 digits which can take the following values:

First digit : represents the cycle in which the program started; can take values 1 or 2

Second digit : represents the institution in which the program was taken ; can take values from 1 to 4

Third digit is always zero

Fourth digit : represents the program taken; can take values from 1 to 3.

This identifier does not exist on the cycle 1 data file. If you need to retrieve a variable from the cycle 1 file you will need to refer to the table on Appendix B which indicates which variables belong to which institution and which program. For example you need to find question hq21, if the PSPLNGID is 1301 then variable H21c1, is associated with institution 3 program 1

5.1.4 Postsecondary Engagement Roster Codebook

Module Name and identifier	Derived variable	Description	Universe
Entry	RecordID	Respondent identification number	All respondents
Postsecondary Engagement – Module K	HPDPSD2	This number given to the Average number of hours of paid work per week during the first year of postsecondary school	Respondents whose 1 st and/or 2 nd post-secondary experience occurred in cycle 2, as well as those whose 1 st or 2 nd post-secondary experience occurred in cycle 1 for which module K questions need to be asked.
	HUWPSD2	Average number of hours of unpaid work in family's business or farm per week during the first year of postsecondary school	(as above)
	KEXPIDD2	Indicates if it is the respondent's 1 st or 2 nd post-secondary experience, in Quebec or non Quebec, in a cegep or non-cegep institution.	(as above)
	KINSTD2	Post-secondary institution identifier which identifies which institution this experience is associated with.	(as above)

The two derived variables KEXPIDD2 and KINSTD2, provides information on the type of postsecondary experience: first or second, in Quebec or not in Quebec, a cegep experience or a non-cegep experience. It also provides a link between the postsecondary experience and the institution to which that experience is associated.

For cycle 2 and beyond, KINSTD2 (postsecondary experience institution identifier) matches PSILNGID, which is the postsecondary institution longitudinal identifier.

If the first postsecondary experience was in cycle 1, the link cannot be established with PSILNGID (which doesn't exist in cycle 1). This identifier does not exist on the cycle 1 data file. If you need to retrieve a variable from the cycle 1 file you will need to refer to the table on Appendix B which indicates which variables belong to which institution. For example if you need question H8, and KINSTID=12 then variable H8b, is associated with institution 2.

In cycle 1, information on postsecondary experiences was not rostered. Appendix B is a table including all variables names (including derived variables) at the institution level, the program level and the person level for cycle 1.

5.1.5 Job Roster Codebook

Module Name and identifier	Derived variable	Description	Universe
Entry	RecordID	Respondent identification number	All respondents
	P1JOBID	Unique job identifier, indicates the position where data in this cycle for this job were collected	Respondents who worked at an eligible job between January 2000 and December 2001
	P1UNID	Longitudinal job identifier which permits following a job across cycles	Respondents who had a job in December 1999
Work Related Questions – Module P1	OJOB2	Cohort B respondents were asked details about chronological order of jobs	Respondents who had at least one job between January 2000 and December 2001
	TENURED2	Total number of months respondent employed at job (regardless of unpaid leaves)	Respondents who had at least one job between January 2000 and December 2001
	TNUR01D2	Total number of months in 2000-2001 respondent employed at job (regardless of unpaid leaves)	(as above)
	JOBOCCD2	4 digit occupation code (SOC 1991) for eligible jobs	(as above)
	JOBOCRD2	2 digit occupation code (SOC 1991) for eligible jobs	(as above)
	JOBINDD2	4 digit industry code (NAICS 1997) for eligible jobs	(as above)
	JOBINRD2	2 digit industry code (NAICS 1997) for eligible jobs	(as above)
	JSDATD2	Start date of the job	(as above)
	WKST01D2 to WKST24D2	Working at job status - Flag for each month in 2000-2001 indicating if respondent was employed and working at the job - for jobs 1 to 7	(as above)
	WKFTPTD2	Full-time/part-time status for respondent who was working at a job in December 2001	(as above)

5.1.6 Job Details Roster Codebook

Module Name and identifier	Derived variable	Description	Universe
Entry	RecordID	Respondent identification number	All respondents
	P1JOBID	Unique job identifier, indicates the position where data in this cycle for this job were collected	Respondents who worked at an eligible job between January 2000 and December 2001
Employment – Module P2	EPHS12	Earnings per hour when first started job	Respondents who had a job at any time between January 2000 and December 2001 and who were paid employees or self-employed when first started this job
	EPWSI2	Earnings per week when first started job	(as above)
	EPMSI2	Earnings per month when first started job	(as above)
	EPHEI2	Earnings per hour when last worked at job	(as above)
	EPWEI2	Earnings per week when last worked at job	(as above)
	EPMEI2	Earnings per month when last worked at job	(as above)
	NHWPMSI2	Number of hours usually worked per month when first started working at job	Respondents who were employed at a job between January 2000 and December 2001
	NHWPMEI2	Number of hours usually worked per month when last worked at job	(as above)
	NWWPMSI2	Number of weeks usually worked per month when first started at job	(as above)
	NWWPMEI2	Number of weeks usually worked per month when last worked at job	(as above)
	HWSD2	Indicates whether the respondent usually worked 30 or more hours per week when first started working at job	Respondents who had at least one job between January 2000 and December 2001
	HWED2	Indicates whether the respondent usually worked 30 or more hours per week when last worked at job	(as above)
	NMWO1D2	Number of months in 2000-2001 where respondent did some work at job (i.e. Total months employed at job less number of months respondent had unpaid leaves, if there were any)	Respondents who were employed at a job between January 2000 and December 2001

6.0 YITS Scales

In YITS cycle 2, two of the concepts - high school engagement and social support were measured through the use of a scale. The high school engagement scores were calculated for members of the 15 year old Reading Cohort who attended a high school, junior high or elementary school between September 2000 and December 2001. Social support was assessed for all members of the 15 year old Reading cohort. For more theoretical details about any of the mathematical/statistical concepts discussed in this chapter, please see the Statistics Canada technical document *Analysis of Scales for YITS Cycle 1 Main Survey* that should be available in the fall of 2005.

Although there were some 18-20 year old cohort respondents who were eligible to receive the questions related to the high school engagement scale, the YITS team decided that there were too few of them (less than 10 per cent of all the 18-20 year old cohort cycle two respondents) to create a scale. It was felt that those students were a special population and that any scale scores created for them would not be representative of the larger population.

This chapter is divided into six sections:

- **6.1 Defining Scales and Their Uses:** introduces the user to a scale, provides a justification for scaling and describes the type of scaling applied within YITS.
- **6.2 Scale Development:** discusses the theoretical procedure that the YITS team used to form the scales.
- **6.3 The High School Engagement Scale:** introduces the user to the concept of school engagement as well as to the High School Engagement Scale and its subscales and sub-subscales. This section also provides justification for the creation of a High School Engagement Scale.
- **6.4 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 section 6.2 to the theoretical model discussed in section 6.3. The results of the analysis are broken into several components: testing model validity, estimating scores, testing scale reliability and testing scale validity.
- **6.5: Social Support Scale:** Provides the results and analysis of an adapted version of the social support scale.
- **6.6 References:** contains a list of the references cited or consulted within the previous five sections.

6.1 Defining Scales and Their Uses

6.1.1 What is a Scale?

For the purposes of social science research, a scale is a type of composite measure consisting of several items (questions) 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. The term scale is also used within the context of this chapter to refer to the theoretical concept upon which the scales are derived.

6.1.2 Why Use a Scale?

The use of scales in data analysis allows researchers to estimate a measure of a particular underlying (latent) concept when the items measuring the concept 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 relationships that may exist among concepts.

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 responses to a group of survey questions (items). 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 (see section 6.2.4) than do individual items. Indeed, a scale increases in reliability as the number of items contained within it increases. An item can contain information about the construct being measured (signal) and confounding variance due to measurement error and information uniquely associated with that item (noise). Using a scale helps to reduce the effects of noise and increases the amount of information available for analysis. Therefore, a multiple-item scale provides more information to analysts than does a single item. 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.

6.1.3 What Type of Scales are Used in YITS?

All of the scales used in cycle 1 and in cycle 2 of YITS are modeled after the Likert Scale (Likert, 1932). This type of scale is valued for the ordinality of its multiple response categories. This allows researchers to compare the relative strength of agreement of survey respondents to any particular item. For example, a particular question with four categories may require respondents to express their views on an issue from four ordinal scale values such as 1 (strongly disagree), 2 (disagree), 3 (agree) and 4 (strongly agree).

A Likert-type scale is a highly reliable tool for rank-ordering people when attempting to measure attitudes or opinions that they hold on a topic. 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 given to them. 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 an 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. Scales need not have the same number of item categories for all of the items in the scale.

In many of the items that make up the various YITS scales, an item category representing a neutral response such as “*neither agree nor disagree*” or “*sometimes*” has been excluded from the categories available to the respondent. 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 variability in responses to each item and therefore 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, with respect to an item, even if these tendencies 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 he or she should indicate a response of “*don't know*”. Other experts, however, think that the neutral response category is necessary for those respondents who truly do not have an opinion on the statement presented to them. The theory and design of survey scale items is discussed in more detail in *Survey Research Methods* (Fowler, 1995).

If one chooses to use a previously constructed scale in a survey, then it is important to consider whether this scale addresses the issues that the survey is attempting to measure. This is especially important in research domains where there are debates within the social science literature as to what the concept should measure. One such discussion occurs when there are many different scales that have varying degrees of specificity and/or focus upon particular aspects of a concept. Although differences between global measures towards a concept and measures of specific attitudes or facets of a concept are often overlooked in research, their measures and their behaviours are not necessarily equivalent (Rosenberg et al., 1995).

6.1.4 Response Biases

The systematic tendency for participants to respond to rating items independent of their content (what the item is designed to measure) is referred to as response bias. This tendency is also referred to as a response set or as a response style depending upon the context. A response set is a temporary reaction to a situational demand. These demands can include time pressure or expected public disclosure. Bias could also result from context issues such as the format of the item or the nature of previous items in the questionnaire. If an individual displays bias consistently over time and situations, then this bias is regarded as his or her response style (Paulhus, 1991, p. 17).

6.1.5 Negative and Positive Questions

It is recommended that both negatively and positively worded questions be included in widely-used rating scales within the discipline of psychology in order to reduce a variety of response biases including acquiescence³ (Marsh, 1996). This is done under the working assumption that positively and negatively worded items represent the same concept.

Sometimes, however, factor analytic techniques (section 6.2.1) indicate different separate factors resulting from the positive and negative worded questions. The crucial question that must be answered in these cases is whether this distinction between the positively and negatively worded item factors is in fact substantively meaningful. Alternatively, it is possible that this distinction is merely an artefact of a person's response style (Paulhus, 1991, p. 48). These two explanations have quite different implications; however, distinguishing between them can be difficult (Marsh, 1996).

6.2 Scale Development

To gain a better understanding of the school engagement process, the YITS team designed a school engagement scale. It should be noted that the school engagement scale developed for YITS may be a different construct than that which has been applied by PISA in cycle one. Documentation on PISA's high school engagement model and international comparisons of school engagement using PISA's model can be found at www.pisa.oecd.org.

Many studies have attempted to define and estimate the rates of school leavers and determine variables that correlate with school-leaving. Furthermore, various reports have been written that detail intervention strategies to encourage school leavers to return or that encourage older students to remain in school. However, few studies are based upon a systematic understanding of the development process that leads to a student's withdrawal from school (Finn, 1989).

There were three main steps in the estimation of the scale scores. The first step was to use linear factor analysis to investigate if the underlying theoretical structure of items and subscales was supported by the data (section 6.2.1). 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 6.2.2). Finally, reliability and validity checks were performed on the estimated scores (section 6.2.3 and section 6.2.4). The general procedures that were adopted for each of these steps are described below.

³ Acquiescence is the tendency to agree rather than disagree with item statements (Paulhus, 1991, p. 46). Some individuals referred to as "yea-sayer", tend to agree with statements whereas other individuals referred to as "naysayers" tend to disagree with statements.

6.2.1 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. Factor analysis is also one way to help achieve the goals of an item analysis. Item analysis is the verification that items are related to their proposed constructs and that the strength of these relationships are adequate for measurement purposes (Gorsuch, 1997).

According to Comrey and Lee, the goal of a 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, not all sets of variables are equally good in representing this relationship. Moreover, there is not necessarily only one “correct” concept or “real” factor for a given domain of interest (Comrey and Lee, 1992, p. 245).

The relationship of each variable to each of the factors, referred to as the loading on a factor, provides a way for the analyst to quantitatively assess how an item interacts with other items. The strength of these loadings on different factors indicates to the analyst whether an item is related to none of the proposed factors, to only one of the factors or to more than one factor (multiple loadings). The greater the level of loading of a variable on a factor, the greater the amount of overlapping (common) variance⁴ between a data variable and a factor and the more an item is a pure measure of this factor.

Items most clearly associated with only one factor can become part of a scale for the construct underlying that factor. Items not strongly associated with any of the factors can be dropped from further analysis. Items can be strongly associated with more than one factor. These items may indicate a relationship between the factors on which the loadings occurred, in which case many items should load on more than one factor and the multiple loadings can be explained mathematically. Alternately, they may indicate the presence of an unknown or confounding factor that is related to the proposed factors. When this is the 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 “significant”. Although no formal test to determine significance has been developed, Comrey and Lee provide a benchmark to use in interpreting variable-factor correlations⁵ (Comrey and Lee, 1992, p. 243). They base their benchmark upon an examination of the percentage of variance of the item common to the factor (See Table 6.1).

Table 6.1 Scale of Variable-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 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.

⁴ 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 a scale.

⁵ Orthogonal factor loadings or structure coefficients

Factor Analysis: Procedures

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, 1995, p.74). In this case, the bias may affect the responses due to imprecise translations. This problem could create potentially different interpretations of the question. This in turn may lead to different response patterns on a question between those respondents who were administered the questionnaire in English and those who were administered it in French.

To examine the possibility of translation bias, the data was first divided up into two samples according to the language of questionnaire administration: English and French. Comparison of the results from each linguistic sample was undertaken. Greater dissimilarities between the results would 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 was regarded 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 half-sample to determine how many factors should be extracted from the data. Factor loadings of each variable were compared between the half-samples and across language. Loadings were examined under a rotation of the factor loadings 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 interpreted 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.

For every factor analysis presented within this document, the value of the items for each respondent was multiplied by the respondent's normalized weight in order to obtain a design-consistent estimate of the variance-covariance matrix. A normalized weight was calculated by dividing a respondent's cycle 2 survey weight (see Sections 9.0 through 10.0 Y for a discussion on weighting in YITS) by the average weight of all eligible respondents in the sample. Thus, in theory, the sum of the normalized weights is equal to the sample size of the eligible respondents.

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, 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 $k + 1((k+1) - \text{item value})$. 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 6.2.3** for a discussion of Cronbach's alpha).

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 of these items.

6.2.2 Estimation of the Scores

The results from the factor analysis were used to determine what items loaded onto each factor (**section 6.2.1**). Once the factor structures were analysed 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, factor scores, a linear composition based upon the linear factor loadings developed under the factor analysis framework (the standard classical method) and non-linear maximum likelihood estimation based upon item response theory (IRT). Both methods were examined using normalized survey weights. Weights were incorporated into the analysis processes in order to derive design-consistent point estimates of the scores.

Factor analysis requires that the scale test data have the property of interval or ratio data. However, some people argue that the test data only have the properties of ordinal data. Whether psychological test scores should be considered ordinal or interval data is a subject of some debate within the social science community (Crocker and Algina, 1986, pp. 61-63). Generally, it is agreed that if it can be demonstrated that the scores produce more information that is useful when they are treated as interval data, then they should be treated as such. On the other hand, if treating the data as interval level measurement does not improve, or actually decreases, their usefulness, then only their rank-order information should be used (Crocker and Algina, 1986, p. 61).

IRT is able to control better for the ordinal nature of the data than is factor analysis. The software program PARSCALE⁶ (based upon the theory developed by Eiji Muraki and Darrell Bock) was applied to calculate the IRT scores and the estimates of the score's measurement errors. Scores released for YITS scales were based upon a parametric IRT approach.

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 (categorical) data (Muraki, 1992). With this particular extension, one obtains an overall slope parameter for each item and a difficulty parameter for each category within an item. The YITS team adopted a particular form of this model where the difficulty parameter is split into a category parameter (a common parameter to all items within a specific block) and an item-specific location parameter.

For estimating IRT scores, the population distribution of the scores was specified to have a mean of zero and a standard deviation of one. Once standardized, the respondent's estimated score, in this case, can be interpreted as the number of standard deviations of the population of interest above (if positive) or below (if negative) the mean.

A response pattern of mostly extreme positives (i.e., strongly agree to all positively worded questions and strongly disagree to all negatively worded questions) is most likely to have been produced by an individual with a highly positive standard score. Conversely, a response pattern 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 who has a score on the scale of around zero. A response pattern that results from choosing the option that is the most extreme, in most cases, may be due to an internal bias by the respondent towards extreme responses or it may indicate a strong opinion by the respondent on the subject of inquiry (Paulhus, 1991, p. 49).

6.2.3 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.

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

⁶For more information about PARSCALE, please see its user guide. (Du Toit, 2003) .

values range from zero to one. Cronbach's alpha can be regarded as a lower bound on reliability; the true reliability of the scale is at least as high as the value of reliability calculated using this measure.

One common misconception about Cronbach's alpha is that a relatively high alpha value for a factor indicates that the administered test items are one-dimensional (the items are associated with only one common underlying factor). Indeed, as “[Cronbach's] alpha is a function of item covariance, and high covariance between items can be the result of more than one common factor, [Cronbach's] alpha should not be interpreted as a measure of the test's one-dimensionality” (Crocker and Algina, 1986, p. 142).

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_j , 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, 1986, p. 352).

$$\eta_j = \sqrt{1 - \frac{\sigma^2_{E_j}}{\sigma^2_j}} \quad (6.1).$$

For a given scale j , $\sigma^2_{E_j}$ 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.

6.2.4 Testing 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. The validation process checks to see if the construct appears to be the same as it is commonly defined. This is despite modifications that may be made to the number or wording of items in the scale for operational constraints. These tests involve evaluating the proposed scales or subscales by comparing their estimated scores on the scales to scores on other relevant scales or to the values of other relevant variables (criterion validity). They can also involve the comparison of different identifiable groups of respondents on the scale of interest (known-group validity). It can also be important to show that a scale does not have high correlation with attitudes that it is not designed to measure (discriminant validity). 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 or not related to the scale of interest and the form that this relationship is expected to take. Within the YITS 15-year-old Reading cohort, criterion and known group validity tests were performed on the scales.

6.3 The High School Engagement Scale

6.3.1 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 (Voelkl, 1996). A student's engagement, or involvement with school, has been linked to academic achievement and to the probability of the student 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 [and] juvenile delinquency" (Voelkl, 1996). Research indicates that children and youth who are engaged with school are more likely to achieve success and become school graduates (Voelkl, 1995, 1997). Nevertheless, there have been very few efforts to define and study the concepts of school engagement formally (Finn, 1993).

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. "[Youth] who have the needed skills and knowledge will be better equipped to succeed as global trade expands and as 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" ((Bowly and McMullen., 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, it is important to identify and try to understand the factors that influence a youth's engagement in school.

6.3.2 What is the YITS High School Engagement Scale?

To gain a better understanding of the school engagement process, the YITS team designed a school engagement scale. School engagement is an important issue likely starting in a student's early school years and continuing throughout a student's entire academic career (Finn, 1993); (Voelkl, 1996). The High School Engagement Scale administered to the YITS 15-year-old Reading Cohort in cycle two attempted to assess engagement for the student's last year in school. The engagement questions focused upon this year since it was believed 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. Furthermore, since school engagement questions had already been asked of these students in cycle one, the earlier timeframe had already been covered. It should also be observed that although there were some 17-year-old student respondents in the sample for whom their current grade was a grade less than high school⁷, the vast majority of the sampled students (approximately 98 %) were in high school.

Two sets of attributes may place a student at risk for educational difficulties-status risk factors and behavioural risk factors. Status risk factors are demographical and historical characteristics of the students whereas behavioural risk factors are a series of actions and activities that, if not manifested by the youth, will decrease the likelihood that successful school outcomes will be achieved (Finn, 1993). While both of these types of risk factors are important, the YITS High School Engagement Scale focuses upon examining behavioural factors

The High School Engagement scale is composed of the Academic Engagement and Social Engagement subscales. The Academic Engagement subscale is further divided up into participation and identification sub-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.

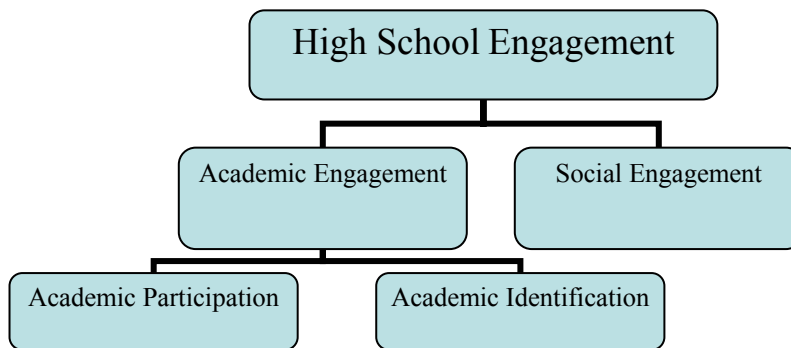
⁷In this chapter, by high school the YITS team means grades 9- 12 or the equivalent (regardless of province)

For the YITS 15-year-old Reading Cohort and for the YITS Mathematics Cohort, an adapted version of items developed by Jeremy Finn 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 cover academic and social identification. She refers to these concepts as school identification (Voelkl, 1996). To these scales, a few additional items were added⁸ to more adequately represent the concepts proposed in the theoretical model. Using these items, the YITS team adapted the model and definitions of school engagement provided in Norris et al (1998).

6.3.3 Structural Model of High School Engagement

A theoretical construction of the School Engagement Model follows below.

Figure 1 Model of High School Engagement for the YITS Reading Cohort (15-year-old)



6.3.4 Description of High School Engagement

Academic Engagement

A student’s academic engagement is defined as their 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. Academic Engagement is broken down into two parts: Academic Participation and Academic Identification.

Academic Participation

Finn (1989) identifies a four-part taxonomy to academic participation. The first three levels are the focus of YITS. The first and most basic level of this classification system is the acquiescence to the need to attend school, to be prepared and to respond to directions and questions. This level is all that is expected of students in the primary grades. Yet, it remains important throughout a student’s career. However, even some level one-participation behaviours will be resisted by some students.

Level two-participation behaviour involves students demonstrating initiative-taking behaviours. These actions can include a student asking questions of and having a dialogue with the teacher as well as a student displaying enthusiasm for school by spending extra time in the classroom before, during or after school or by doing more class work or homework than is required. This enthusiasm may eventually translate into participation in school-related clubs, and into community activities such as science fairs and internships. For those students who experience academic difficulty, “help seeking behaviour” is regarded as an important set of

⁸ 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.

initiative-taking behaviours. The decision by a student to obtain help is affected by both his/her recognition of a learning difficulty and his/her desire to overcome it. It is also influenced by a student's perception of the response of the helper and other classmates to help-seeking behaviour on other occasions.

The third level of behaviours involves participation in the social, extracurricular, and athletic aspects of school life in addition to or as a substitute for extensive participation in academic work.

While some work cited by Voelkl (1995) has shown that there is a weak relationship between school warmth (how students perceive their academic environment) and academic achievement, recent work performed by Voelkl has found this relationship to be insignificant after the effects of participation were eliminated (Voelkl, 1995). Thus, while a warm and supportive classroom environment is important to a student's development, this importance may result from creating an environment where participation is encouraged.

The failure to participate in classroom activities and the expression of anti-participatory behaviours are collectively referred to as non-participation. Finn (1993) cites many studies that have found links between participation and school performance. Non-participatory behaviours can be classified into two forms a) failing to take advantage of constructive strategies for learning, b) engaging in negative behaviours that impede learning.

Academic Identification

Identification is used by social scientists in various ways. Nevertheless, the concept always has as its core the idea of an "affinity target" (a subject of attachment). While the term identification often signifies attachment with another person, it can also be used to describe the relationship that exists between a person and an institution such as a school. "A person comes to identify with a place or structure that represents certain expectations, values, beliefs, and practices" (Voelkl, 1997). During a youth's childhood and adolescent years, when schooling has a predominant role in his or her everyday life, school may become central to a youth's self-view.

Personal adjustment by the individual to the institution can be reflected in the attitudes that a student demonstrates towards the school such as academic motivation, comfort in school and the exhibition of acceptable or unacceptable personal and social behaviours in the classroom. Students who have a higher identification with school are expected to have a higher level of academic achievement and a greater persistence throughout the grades. By contrast, those students who do not identify with school may be predisposed to a pattern of negative school behaviour and to eventual withdrawal from school (Voelkl, 1997).

Within the YITS Reading Cohort, the focus of attention is on two components of identification, valuing and belonging. Belonging refers both to a sense of fit between the students' perceived needs and the offerings of the school and to a student's perception that they are in an environment where they are cared about and respected. Researchers suggest that a sense of belonging to a school increases the expectations of success for a student through enhancing the idea that one has the necessary skills and the available social resources to overcome difficulties and succeed in school. It also increases a student's belief in the value of their academic work (Goodenow, 1993). Valuing refers to a student's appreciation of success in school-relevant goals and to a student's endorsement of the goals of education in general⁹ and of academics in particular (Voelkl, 1997).

A student who fails to identify with school is expected to have a lack of valuing for the school and a lack of feelings of belonging to the school. Voelkl (1996) characterizes the emotional state represented by a student who disidentifies with school as follows:

The student [who] does not feel a sense of belonging or of valuing school, does not feel like an accepted member, has little or no sense of 'fitting in', does not feel comfortable or adequate, fails to incorporate school into his or her self-definition, may feel anger or hostility towards school, would rather be in a setting other than school, has little feeling of commitment to school, and may be distrustful and suspicious of the institution and those who represent it.

While belonging, or its absence, likely influences motivation and participation in school for most adolescents, its effects should be especially important for those students at the low end of the identification scores and less important for those students at the high end. This is because students at the low end are at greater risk of becoming school leavers (Voelkl, 1997).

⁹ These goals are the value of school as a social institution and as a tool for personal advancement.

Academic Participation and Academic Identification Relationship

Academic identification can be regarded as the extent to which a student has bonded with his or her school and substantively amalgamated the connection with this institution into his or her self-concept and lifestyle. Students' endorsement of the academic environment, indicated through their feelings of identification and participation in the classroom, is part of a cycle that affects academic achievement. This is a behaviour-attitude relationship. "In particular, students who develop a sense of identification with school are likely to experience greater education gains, expressly as a consequence of active participation in the classroom and in school generally" (Voelkl, 1996). By contrast, it has been found that student's lack of identification in school is "partially a function of the degree of academic success experienced in school and partially a function of the degree to which [youth] participate in classroom learning activities" (Voelkl, 1997). Thus, academic identification is a fundamental part of the process through which academic participation manifests itself.

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. This connection reflects the extent to which students feel personally accepted, respected, included and supported by others in the school's social environment.

Social Engagement is influenced by a student's personal traits as well as situational factors within the school environment. 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. YITS main focus was with the belonging aspect of social engagement.

6.4 Analysis of High School Engagement Scale

The main steps in the development of the scales presented in **section 6.2** were applied to the high school engagement model illustrated in **section 6.3.3** and discussed in **section 6.3.4**. Analysis of the YITS 15-year-old Reading Cohort data revealed that the data supported three factors: Academic Participation (AP), Academic Identification (AI) and Social Engagement (SE). Seven items loaded on the Academic Participation sub-subscale, eighteen items loaded on the Academic Identification sub-subscale and nine items loaded on the Social Engagement subscale.

The empirical factor structure of the High School Engagement Scale for the 15-year-old YITS Reading cohort between cycle one and cycle two is the same. However, there is a difference in the timeframe focused upon in the two cycles. The school engagement questions asked in cycle one focused on the respondent's current school year, whereas in cycle two, they focused upon the respondent's last year in formal schooling at either the elementary or the secondary school levels. This last year may not necessarily have been the year in which the respondents were interviewed.

The method of collection differed from cycle one to cycle two. In cycle one, the 15-year-old YITS Reading cohort respondents were administered a self-complete questionnaire in school whereas in cycle two, the data was collected via a CATI process. There are some differences in the academic participation sub-subscale between the two cycles. While this subscale in cycle one focussed mainly upon homework related questions, the sub-subscale in cycle two had a broader focus. Researchers are cautioned to be aware of these changes when comparing results between the two cycles.

It should be noted that while the respondents are now close to the age of the 18-20 year old cohort respondents in cycle one, the empirical structure for the high school engagement scale is different between the two cohorts. High school engagement for the 18-20 year old cohort had a two-factor concept – academic engagement and social engagement while the concept for the 15-year-old YITS Reading cohort is a three-factor concept.

6.4.1 Testing Model Validity

The 35 items that were hypothesized to define the concept of high school engagement were administered to a sample of 26, 693, 17-year-olds (respondents in cycle 1 of YITS 2000). All 17 year-old respondents who took elementary school, junior high or high school courses anytime during the period between September 2000 and December 2001, excluding those who had never attended an educational institution (i.e. had done their studies completely through home schooling or through distance education) were considered to have been eligible to have received high school engagement related questions. Twenty cycle two respondents who did receive a scale score in cycle one were not administered the scale questions in cycle two as they indicated that they had never attended a formal educational institution (home-schooled).

The 15-year-old cohort 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 three-factor model – Academic Participation (YSHACPS2), Academic Identification (YSHACIS2) and Social Engagement (YSHSOES2). – having seven items, eighteen items and nine items respectively.

Finn (1993) proposed that his academic identification sub-subscale broke into two components - belonging and valuing. Voelkl (1995), when using similar questions to those of Finn (1993), also found evidence of a split into two components. However, on comparing the two-factor solution (with valuing and belonging as separate dimensions) and the one factor solution (identification), Voelkl concluded that the one factor solution fit the data just as well as the two-factor model did. Using the data collected in cycle two of the YITS Reading Cohort data, the YITS team was unable to find a clear split for some of the items between the belonging and valuing components of identification. For this reason, and for consistency with the cycle one framework (see cycle one User Guide), the YITS team decided to retain only the overall academic identification sub-subscale.

Items Dropped from Further Analysis

Altogether, three items were dropped from the academic identification, academic participation sub-subscales and the social engagement subscale for reasons discussed below. If an item had a factor loading above 0.40, it was considered to load strongly enough to keep it in the final scale. When an item had loadings that were between 0.32 and 0.40, the item was kept if the YITS team felt that it contributed strongly to the concept being measured.

Some items that were present in cycle one were no longer asked of the respondents in cycle two. In particular, PISA questions dealing with the number of hours doing subject specific homework were dropped. Similarly, some of the questions that were dropped from the scales in cycle one for various reasons (see cycle 1 User Guide) were not asked again in cycle two. However, there were some questions that were dropped from cycle one scales that were asked again in cycle two. These questions were re-evaluated to determine whether they would now load appropriately (i.e. on the expected subscale) and with sufficient loading to justify retaining them for the cycle two sub-scale.

In cycle two, students who indicated in question F2Q12 that they did zero hours of homework, were then asked question F2Q12a, “Were you assigned any homework?” In order to keep the analysis similar to that applied in cycle one, it was decided to take an unconditional approach to responses provided to F2Q12. Thus even if the respondent indicated in F2Q12a that they did not have any homework to do outside of class they were still considered to be in scope for question F2Q12 and were classified as part of those respondents who indicated they did zero hours of homework for the purposes of creating the scales.

Table 6.2 Academic Identification Items Description and Loadings

Item Code Codebook	Item Description	Reason for Dropping Variable
F2Q19	When school work was very difficult, I stopped trying	Did not have strong loading on academic participation as well, the communality of the variable was low. This suggests that little of the overall variance is explained by its relationship to the factor.
F2Q31	There were teachers or other adults in my school whom I could talk to if I had a problem	Similar to cycle one, this item did not load on social engagement and as the location of its loading did not correspond to the expectations for this item by the YITS team, it was dropped. A preliminary investigation indicated that this item did load on academic identification.
F2Q56B	Number of hours spent participating in school clubs, teams or other school organizations	Similar to cycle one, it did not load on Social Engagement as expected and loaded weakly on all factors. It is possible that had there been more social engagement questions related to social participation that this question would have loaded on a social participation dimension.

Items Retained

It was observed that there was a large difference in the average number of hours of homework reported by French and English respondents. Further investigation of this difference determined that this was because French Quebec students reported less homework than did students in the rest of the country including their English counterparts in Quebec¹⁰. This large difference in the mean was not seen in cycle one. The reason for this difference requires further research.

To prevent generating potentially biased item parameter estimates, the YITS team decided to create two artificial items out of the question F2Q12. One item (F2Q12FQ) was assumed to have been administered to only the French Quebec students and the other item (F2Q12) was assumed to have been administered to all other in scope respondents.

For those respondents whose province of education was missing (either not known, or not stated), their province of residence was used as a proxy in determining whether they received the item F2Q12FQ or F2Q12. If they resided in the province of Quebec, and were given the interview in French, then they were assumed to have been educated in French and were thus assumed to have been administered F2Q12FQ¹¹.

The items retained for each factor, their description and their factor loadings are provided in the following tables.

¹⁰ By English and French Quebec students, it is meant those respondents who were educated in the province of Quebec and who responded to the questionnaire in English or French respectively.

¹¹ There were 26 respondents who did not provide the province in which they were educated to the interviewer of which 16 were assigned to the question F2Q12FQ.

Table 6.3 Academic Participation Items Description and Loadings¹²

Item Code Codebook	Item Description	Factor Loadings (Others)	Factor Loadings (French Quebec)
F2Q12	Hours on homework outside of class.	0.60	0.63
F2Q13	Number of times I skipped class.	0.60	0.53
F2Q14	I paid attention to the teacher	0.60	0.59
F2Q15	I completed my assignments.	0.73	0.73
F2Q18	I did more than the required work	0.56	0.56
F2Q20	When school work was very difficult, I stopped trying	0.69	0.66
F2Q50	I completed my homework on time.	0.76	0.74

Table 6.4 Academic Identification Items Description and Loadings

Item Code Codebook	Item Description	Loadings
F2Q22	I got along with teachers.	0.51
F2Q23	I was interested in what I am learning in class.	0.55
F2Q24	I felt proud to be part of my school	0.58
F2Q25	I was treated with as much respect as other students in my class	0.56
F2Q27	School was one of the most important things in my life.	0.55
F2Q28	Many of the things we learned in class were useless.	0.57
F2Q29	Most of my teachers did not really care about me.	0.65
F2Q30	Most of the time, I would have liked to have been any place other than school.	0.65
F2Q32	Most of what I was learning in school would be useful when I get a job.	0.57
F2Q34	School was often a waste of time.	0.68
F2Q35	School is more important than most people think.	0.50
F2Q36	Most of my teachers did a good job at teaching.	0.62
F2Q40	Most of my teachers really listened to what I have to say.	0.66
F2Q41	If I needed extra help, I received it from my teachers.	0.57
F2Q42	Most of my teachers treated me fairly.	0.64
F2Q48	My school is a place I do not want to go.	0.70
F2Q49	My school is place where I often felt bored	0.66
F2Q55	I was given interesting homework.	0.51

¹² The loadings presented for all following factors presented within this chapter, unless otherwise stated, are from the combined factor analysis on all eligible respondents.

Table 6.5 Social Engagement Items Description and Loadings

Item Code Codebook	Item Description	Factor Loadings
F2Q33	People at school are interested in what I have to say.	0.54
F2Q37	I have friends at school whom I can talk to about personal things.	0.61
F2Q38	I have friends at school who can help me with school work, if needed.	0.59
F2Q43	My school is a place where I feel like an outsider (or left out of things).	0.73
F2Q43B	My school is a place where I make friends easily.	0.77
F2Q44	My school is a place where I feel like I belong.	0.76
F2Q45	My school is a place where I feel awkward and out of place.	0.79
F2Q46	My school is a place where the students seem to like me.	0.72
F2Q47	My school is a place where I feel lonely.	0.76

The factor loading for the items on all three subscales were similar in magnitude. According to the Comrey and Lee benchmark for rating scale loadings (**section 6.2.1**); the loadings for the Social Engagement subscale with values from 0.54 to 0.79, range from fair to excellent. The loadings for the AP sub-subscale, with values from 0.56 to 0.76, range from good to excellent for the majority of the respondents and between 0.53 and 0.74 for the French Quebecers. The loadings for the Academic Identification sub-subscale, with values from 0.50 to 0.70, range from fair to very good.

6.4.2 Estimating Scores

Scores on the two subscales of Academic Engagement (Academic Participation and Academic Identification) and the Social Engagement score were estimated using item response theory by the program PARSCALE. An estimated score was obtained for a sub-scale or sub-subscale by all eligible respondents who responded to at least one item for that subscale.

In order to obtain good item parameter estimates, similar items are grouped together into what is known as a block. (Muraki, 1992) For YITS, blocks have been formed by grouping items with the same number of categories together. Blocking items together assumes that items so grouped behave similarly. However, for the Academic Participation subscale, when two items F2Q18 and F2Q20 were placed in the same block, PARSCALE did not converge to obtain a solution. It was noted that the distribution of the participants' responses to the item categories for these two items were relatively different. This suggests that the respondents are interpreting these two questions in different ways. The solution adopted by the YITS team was to create a separate block for estimating parameters of F2Q18.

By taking a simple average of the scores of the AP and AI sub-subscales, an Academic Engagement (YSHACES2) subscale was formed. The Social Engagement (SE) subscale was averaged with the scores from the newly formed Academic Engagement subscale to form an overall measure of high school engagement (YSHSCES2). The measurement standard errors of the scores for each respondent on the AP (YSHACPE2) and AI (YSHACIE2) sub-subscales as well as on the Academic (YSHACEE2) and Social engagement subscales (YSHSOEE2) and the overall High School engagement scale (YSHSCEE2) were also estimated and are included along with their respective scores on the YITS 15-year-old Reading Cohort 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. Overall, 436 cases for the AP sub-subscale, 457 cases for the AI sub-subscale and 492 cases for the SE subscale were not derived for this reason. In all of the cases of missing scores, the scores and the standard error of the scores were assigned a value of 99.99999. . Those respondents who were either in the 18-20 year old cohort or who were in the 15 year old reading cohort and were not eligible to

receive the school engagement questions and the standard error of their scores were assigned a value of 99.99996.

The composite scores of respondents on academic engagement and school engagement are not latent trait estimates in the same sense as the sub-scale scores. Rather, these composites are estimates of latent traits for which one does not have any indicators (or have not assigned indicators). Instead of estimating the value of the latent traits of overall engagement using indicator variables, when composite scores are created, one is predicting the overall engagement scores of students based upon other latent variables that are believed to be highly correlated with overall engagement¹³. The benefit of including these composites is the standardization across analyses done by different researchers that are concerned with examining overall school engagement.

The weakness of using a composite scoring method to obtain higher-order concepts is the *a priori* definition of the relationship of the sub-scales to the composites. Subscales are given equal weighting or are weighted according to standard error of the scores. It is not possible to know the "true" weighting coefficients of the sub-scales for the composite scale without having direct measures of the composites (which, if one had, would mean that one would not need the weights). It is recognized that certain analytical techniques such as structural equation modeling may be used to find weights that are optimal for specific analyses. Analysts familiar with these methods may choose to use the subscale scores, or even to use the separate indicator variables to accomplish this. However, even among analysts who do use structural equation modeling methods, including measurement models in a standard analytic model can be unreasonably complicated. Moreover, the weights estimated using this technique may vary considerably across analyses. The YITS team decided that the limitations discussed above were acceptable, given the increased accessibility to the data and comparability across analyses obtained by creating composite scores and providing access to them on the data file.

There were 139 respondents for academic participation and 30 for the academic identification who obtained an extremely high or an extremely low score for the academic participation and academic identification sub-scales i.e. above 4 or below -4 standard deviations. An initial investigation of the value of the individual's item responses related to these scores has determined that the scores for those respondents do seem to be valid scores. These score values represent extreme responses to questions: An extremely high score would represent an extremely highly engaged student who answers the highest quantity to almost every question. An extremely low score would represent an extremely disengaged student who answers the lowest quantity to almost every question. These quantities do however have large standard errors. This suggests that the value of the latent trait was not measured well for these students.

When either of the two sub-subscales that made up the Academic Engagement subscale were not calculated, the score of the respondent on the Academic Engagement subscale was set to missing. There were 458 cases where this happened. If either the Social Engagement subscale score was not calculated or the Academic Engagement subscale score was set to missing, then the value of the High School Engagement scale score was also set to missing. Four-hundred and ninety-three cases (493) were assigned missing scores for the overall School Engagement Scale.

Setting the composite measures to missing was necessary in these cases because including the estimated score from only one of the two sub-subscales that made up the Academic Engagement subscale would bias the results of secondary analyses. For these students, the Academic Engagement subscale would be wrongly estimated as being either academic participation or academic identification, neither of which is fully conceptually interchangeable with academic engagement. A similar argument holds true for using only one of either the Academic or the Social Engagement subscales to form the overall High-School Engagement scale.

The scale scores from cycle one and cycle two were not put on the same metric; thus, while the scales in both cycles were standardized to have a mean of zero and a standard deviation of one, this standardization process was done separately in the two cycles. This means that one cannot accurately compare whether students' level of individual-level engagement increased or decreased between the cycles. It is however possible to examine changes between cycles in the overall patterns of engagement.

The Item parameters for the subscales are provided below.

¹³ Academic participation and academic identification for academic engagement, and by extension academic engagement and social engagement for school engagement

Table 6.6a

Social Engagement Item Specific Parameters		
Item Code Codebook	Slope Parameter	Location Parameter
F2Q33	1.2026	0.71964
F2Q37	0.88241	1.83976
F2Q38	0.90765	1.51201
F2Q43	1.02688	1.64896
F2Q43B	1.45208	1.26085
F2Q44	1.3282	0.89523
F2Q45	1.33901	1.35972
F2Q46	1.84272	1.10211
F2Q47	1.58066	1.37749

Table 6.6b

Social Engagement Category Parameters			
Category 01	Category 02	Category 03	Category 04
0.00000	1.63741	0.67717	-2.31458

Table 6.7a

Academic Identification Item Specific Parameters			
# of Categories	Item Code Codebook	Slope Parameter	Location Parameter
5	F2Q22	0.37259	-2.49974
5	F2Q23	0.53039	-0.77752
4	F2Q24	0.7602	-1.15242
4	F2Q25	0.74222	-1.48273
4	F2Q27	0.56374	-0.87867
4	F2Q28	0.65387	-0.18659
4	F2Q29	0.9498	-1.44902
4	F2Q30	0.82453	-0.2857
4	F2Q32	0.8006	-0.57846
4	F2Q34	1.01671	-1.07683
4	F2Q35	0.81353	-1.56037
4	F2Q36	1.05543	-1.26159
4	F2Q40	1.08167	-1.1302
4	F2Q41	0.92835	-1.33995
4	F2Q42	1.25353	-1.39625
4	F2Q48	0.91438	-0.98732
4	F2Q49	0.86871	-0.10048
4	F2Q55	0.86871	1.07689

Table 6.7b

Academic Identification Category Parameters (5 Category Block)				
Category 01	Category 02	Category 03	Category 04	Category 05
0.00000	1.53979	1.60340	-0.62031	-2.52288

Table 6.7c

Academic Identification Category Parameters (4 Category Block)			
Category 01	Category 02	Category 03	Category 04
0.00000	2.16903	0.40470	-2.57373

Table 6.8a

Academic Participation Item Specific Parameters			
# of Categories	Item Code Codebook	Slope Parameter	Location Parameter
6	F2Q12	0.33692	0.46922
6	F2q12FQ	0.33996	-0.49501
5	F2Q13	0.23002	-1.51248
5	F2Q14	0.67953	-1.48500
5	F2Q15	0.94780	-1.66007
5	F2Q20	0.33727	0.20317
5 (F2Q18 Block)	F2Q18	0.29156	0.26026
4	F2Q50	1.15687	-0.93891

Table 6.8b

Academic Participation Category Parameters (6 Category Block)					
Category 01	Category 02	Category 03	Category 04	Category 05	Category 06
0.00000	1.65444	3.02807	-0.06784	-1.69521	-2.91945

Table 6.8c

Academic Participation Category Parameters (5 Category Block)				
Category 01	Category 02	Category 03	Category 04	Category 05
0.00000	1.16663	0.86213	-0.24936	-1.77939

Table 6.8d

Academic Participation F2Q18 Category Parameters				
Category 01	Category 02	Category 03	Category 04	Category 05
0.00000	1.31886	1.55389	-0.50546	-2.36730

Table 6.8e

Academic Participation Category Parameters (4 Category Block)			
Category 01	Category 02	Category 03	Category 04
0.00000	1.56695	-0.12334	-1.44361

6.4.3 Scale Reliability

Two common measures of reliability, Cronbach's alpha and the index of reliability were estimated.

Table 6.9 High School Engagement Scale Reliability Measures

High School Engagement subscales	Index of Reliability	Cronbach's alpha
Academic Participation	0.96	0.77
Academic Participation (French Quebec)		0.76
Academic Identification	0.92	0.89
Social Engagement	0.97	0.87
Academic Engagement	0.93	-----
High School Engagement	0.96	-----

The magnitude of the reliability indices of the AP and AI sub-subscales and the SE subscale are very close. Thus, the accuracy of the measurements is about the same for all three subscales. The Cronbach's alpha was calculated separately for the academic participation sub-scale for those respondents who were administered the survey in French and schooled in Quebec and for other respondents. Researchers should use these reliability estimates and the standard errors of the scores provided with the micro data file to determine whether these subscales are reliable enough for their purposes.

6.4.4 Testing Scale Validity

Based upon previous research, the variables listed below are expected to have a relationship with the AP, AI sub-subscales and/or with the SE subscales. The last column provides examples where previous researchers and experts found or proposed similar relationships between the construct and the test items. The direction of the correlation and the strength of the relationship between the variable and the subscale are also provided. For these scales, a medium correlation is considered to be from 0.20 to 0.30, and a high correlation is considered to be 0.30 and above. These correlations indicate that the construct being measured in the YITS questionnaire is similar to the construct as it has been defined by previous research and theory.

Table 6.10 Academic Participation Validation

Item Code Codebook	Item Description	Correlation Strength	Correlation Direction	Reference
C_Q01a	Overall grade average	Medium	Negative	(Finn, 1993) The negative direction of the correlation between results from the fact that the categories in questions related to student's grades were ranked from high grades to low grades
C_Q06	Grade average in math	High	Negative	
C_Q11	Grade average in main language course	Medium	Negative	
F2Q64	How many of your closest friends thought completing high school was very important?	High	Positive	(Voelkl, 1995)

Table 6.11 Academic Identification Validation

Item Code Codebook	Item Description	Correlation Strength	Correlation Direction	Reference
C_Q01a	Overall grade average	Medium	Negative	(Voelkl, 1996)
F2Q03	Students were disciplined fairly	High	Positive	(Voelkl, 1996)
F2Q04	It was difficult to make new friends	Medium	Negative	(Voelkl, 1996)
F2Q05	Violence was a problem	Medium	Negative	(Voelkl, 1997)
F2Q06	People at my school respected others as they were	High	Positive	(Voelkl, 1996)
F2Q08	My school was friendly place	High	Positive	(Voelkl, 1996)
F2Q64	How many of your closest friends thought completing high school important	Medium	Positive	(Voelkl, 1997)
F2Q67	How many of your closest friends planned to further their education or training after high school	Medium	Positive	(Voelkl, 1997)
F2Q78	Times per month using marijuana or hash	Medium	Negative	(Voelkl, 1997)

Table 6.12 Social Engagement Validation

Item Code Codebook	Item Description	Correlation Strength	Correlation Direction	Reference
F2Q04	It was difficult to make new friends at my school	High	Negative	(Voelkl, 1997)
F2Q06	People at my school respected others as they were	Medium	Positive	(Voelkl, 1997)
F2Q08	My school was a friendly place	High	Positive	(Voelkl, 1997)
YSHSUPS2	Cycle Two Social Support Scale	High	Positive	(Cutrona & Russell, 1987)

6.5 Social Support Scale

6.5.1 Description of Social Support Scale

Most conceptualizations of support include the following ideas: emotional sustenance, self-esteem building, provision of information and feedback and tangible assistance (Russell and Cutrona, 1987). A number of different social support measures have been developed. These measures differ widely and on multiple dimensions on how they model social support. These instruments differ on whether they assess: 1) structure or the function of support; 2) subjective or objective support; 3) availability or adequacy of support; 4) individual structures or functions or global indices; 5) several individual structures; 6) the role of people available to provide support or simply whether support is available; or 7) the number of people available to provide support or merely the availability of support (irrespective of the number of people). (Cohen and Syme, 1985, p. 15).

While social support does not have a unique concept or an empirical concept, it is still widely used by researchers. “The term [social support] connotes enough that it has proved fruitful even in the absence of denotation” (House and Kahn, 1985, p. 84). It has been suggested that the reason for this usage is that even without a single concept, it captures a common theme in many seemingly diverse phenomena.

Social support is a concept that may help in the interpretation of the differences observed in people’s responses to common problems. Conditions that create distress in some people do not seem to affect others. Researchers theorize that factors exist that can mediate between difficult life conditions and outcomes. Social support is one of these coping mechanisms. (Pearlin and Schooler, 1978); (Pearlin, 1985, p. 57).

The central goal of the social support model proposed by Carolyn Cutrona and Daniel Russell is to understand the processes through which interpersonal relationships enhance or retain both psychological and physiological well-being. The objective of the measure for YITS was to determine the availability of social supports, via friends, family and other sources for the youth. The social support scale used in YITS is a modified version of the Social Provisions Scale developed by Cutrona and Russell. It was based upon similar modifications to the scale adapted for the Canadian NLSCY (Microdata User Guide (2003))¹⁴

The aspects of social support measured in YITS include three aspects of the original model and are classified under the broad category of assistance-related provisions. They are reliable alliance (the assurance that others can be counted upon for practical help), attachment (emotional closeness) and guidance (advice or information). These aspects are most directly related to problem solving within the context of stress. Two items were proposed to measure each of these aspects for a total of six items. All respondents who were in the YITS 15 Year-old Reading cohort were eligible to receive the social support questions.

6.5.2 Model Validity

No strong differences were found between the factor loadings on the English and the French samples and all of the proposed items were kept. The items that make up the factor, their description and their factor loadings are provided below.

Table 6.13 Social Support Items Description and Loadings

Item Code Codebook	Item Description	Factor Loadings
N2Q11	If something went wrong, no one would help me	0.67
N2Q12	I have family and friends who help me feel safe, secure and happy	0.73
N2Q13	There is someone I trust whom I would turn to for advice if I were having problems	0.77
N2Q14	There is no one I feel comfortable talking about problems with	0.71
N2Q15	There is no one I feel close to	0.81
N2Q16	There are people I can count on in times of trouble	0.80

According to the Comrey and Lee benchmark for rating scale loadings (section 6.2.1); the loadings for the Social Support scale with values from 0.67 to 0.81, range from very good to excellent.

6.5.3 Estimating Scores

This scale scores have the code YSHSUPS2 on the dataset and their standard error has the code YSHSUPE2. Due to a few respondents not answering any of the questions upon which the scale was based, the score could not be estimated for 664 respondents. In all of the cases of missing scores, for this scale, the scores and the standard error of the scores were assigned a value of 99.99999. . Since members of the 18-20 year old YITS cohort were not administered the Social Support Scale, they were all assigned a valid skip value of 99.99996 for the score and the standard error of the scores.

¹⁴ Statistics Canada Microdata User Guide (2003) – *National Longitudinal Survey of Children and Youth – Cycle 4*. Statistics Canada

Item Parameters for the Social Support Scale

Table 6.13a

Social Support Item Specific Parameters		
Item Code Codebook	Slope Parameter	Location Parameter
N2Q11	1.10354	1.55180
N2Q12	1.32307	-1.68189
N2Q13	1.53252	-1.66038
N2Q14	1.03906	-1.29787
N2Q15	1.75363	-1.42650
N2Q16	1.79899	-1.42925

Table 6.13b

Social Support Category Parameters			
Category 01	Category 02	Category 03	Category 04
0.00000	0.52121	0.89417	-1.41538

6.5.4 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 Social Support scale is 0.85 and the value of the index of reliability is 0.96. Researchers should use these reliability estimates and the standard errors of the scores provided with the micro data file to determine whether or not this scale is reliable enough for their purposes.

6.5.5 Testing Scale Validity

With social support scales, it is particularly important to demonstrate that one is actually measuring social support and not some other related personality measure. This is demonstrated by discriminant validity. In particular, one wishes to demonstrate that the support model that is proposed is not highly correlated with such measures as social anxiety, personal competence, and social desirability (Cohen and Syme, 1985, p. 14). Using this scale, Cutrona and Russell (1987) assessed discriminant validity and determined that social support does measure a concept that is distinct from these other measures. Hence, the Social Provisions Scale does contribute to the description of the concept of psychological distress in a way that is different from these related variables. However, it is also necessary as displayed in Table 6.14 to demonstrate that the social support scale used by YITS is correlated to other relevant scales and/or variables (using criterion-based validity).

Based upon previous research, the variables listed below are expected to have a relationship with the Social Support Scale. The last column provides examples where previous researchers and experts have found or proposed similar relationships between the construct and the test items or where the concept being measured in the scale is clearly similar to the concept being measured in the test item. The direction of the correlation and the strength of the relationship between the variable and the scale are also provided. These correlations indicate that the construct being measured in the YITS questionnaire is similar to the construct as it has been defined by previous research and theory. For the Social Support Scale, those items that had a correlation above 0.30 were considered to have a high correlation and those items that had a correlation above 0.20 but below 0.30 were considered to have a medium correlation.

Table 6.14 Social Support Validation

Item Code Codebook	Item Description	Correlation Strength	Correlation Direction	Reference
F2Q04	It was difficult to make new friends	Medium	Negative	(Cutrona & Russell, 1987)
F2Q08	My school was friendly place	Medium	Positive	(Cutrona & Russell, 1987)
F2Q31	There were teachers or other adults in my school whom I could talk to if I had a problem	Medium	Positive	(Cutrona & Russell, 1987)
YSHSUPS2	Cycle Two Social Engagement Subscale	High	Positive	(Cutrona & Russell, 1987)□
YSHACIS1	Cycle Two Academic Identification Subscale	High	Positive	Students who perceive that they belong to school are more likely to feel supported. Goodenow notes that one aspect of specific relevance to education is “student’s sense of belonging in the classroom”, the extent to which student’s feel accepted, respected included and supported by others in the school social environment (Goodenow, 1993).

6.6 Scale References

Bowlby, 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.

Cohen, S. and Syme, S. L. (1985). “Issues in the Study and Application of Social Support”, (pp. 1-22) in *Social Support and Health*. Cohen, S. and Syme, S. L. (Eds.). San Diego, California: Academic Press.

Comrey, A.L. and Lee, H.B. (1992). *A First Course in Factor Analysis*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

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

Cutrona, C.E. and Russell, D.W. (1987). “The Provisions of Social Relationships and Adaptation to Stress”, *Advances in Personal Relationships*, Vol. 1, 37-67.

Du Toit, M., (ed.) 2004, Muraki, E. *IRT from SSI*, Chicago, Illinois: Scientific Software International.

Documentation of the Scales used in the National Longitudinal Survey of Children and Youth, Cycles 1-3, Internal Document. Statistics Canada.

Finn, J.D. (1989). “Withdrawing from School”, *Review of Educational Research*. Vol. 59, No. 2, 117-142.

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

Finn, J.D. and Rock, D.A. (1997). “Academic Success Among Students at Risk for School Failure”, *Journal of Applied Psychology*, Vol. 83, No. 2, 221-234.

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

- Goodenow, C. (1993). "The Psychological Sense of School Membership among Adolescents: Scale Development and Educational Correlates", *Psychology in the Schools*. Vol. 30, 79-90.
- 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, Massachusetts: Allyn and Bacon.
- House, J. L. and Kahn, R. L. (1985). "Measures and Concepts of Social Support" (pp. 83-108). In *Social Support and Health*. Cohen, S. and Syme, S. L. (Eds.). San Diego, California: Academic Press.
- Likert, R. (1932). "A Technique for the Measurement of Attitudes", *Archives of Psychology*. No. 140, 1-55.
- Marsh, H.W. (1996). "Positive and Negative Global Self-Esteem: a Substantively Meaningful Distinction or Artifacts?", *Journal of Personality and Social Psychology*. Vol. 70, No. 4, 810-819.
- Microdata User Guide. (2003). - *National Longitudinal Survey of Children and Youth - Cycle 4*. Statistics Canada.
- Muraki, E. (1992). "A Generalized Partial Credit Model: Application of an EM Algorithm." (Research Reports Educational Testing Services RR-92-06) Princeton, New Jersey: Educational Testing Services.
- 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. (A revised version of this paper entitled "Measuring School Engagement" appeared in June 2003 in *Education Quarterly*)
- Paulhus, D. L. (1991). "Measurement and Control of Response Bias", (pp. 291-372). In *Measures of Personality and Social Psychological Attitudes: Volume 1 of Measures of Social Psychological Attitudes*. Robinson, J., Shaver, P., Wrightsman, L. (Eds.), San Diego, California: Academic Press.
- Pearlin, L.I. and Schooler, C. (1978). "The Structure of Coping", *Journal of Health and Social Behaviour*, Vol. 19, 2-21.
- Pearlin, L.I. (1985). "Social Structure and Processes of Social Support", (pp. 43-60). in *Social Support and Health*. Cohen, S. and Syme, S. L. (Eds.), San Diego, California: Academic Press.
- Rosenberg, M. et al (1995). "Global Self-Esteem and Specific Self-Esteem: Different Concepts, Different Outcomes", *American Sociological Review*. Vol. 60, 141-156.
- Smith, W.J. et al. (1998). *Student Engagement in Learning and School Life: National Project Report*. Montréal: McGill University, Office of Research on Educational Policy.
- Statistics Canada. (2000). *T-00-5E (September 2000) Youth in Transition Survey – Project Overview*. Catalogue no. MP32-30/00 – 5E. Statistics Canada
- Voelkl, K. E. (1995). "School Warmth, Student Participation, and Achievement", *Journal of Experimental Education*. Vol. 63, 127-138.
- Voelkl, K. E. (1996). "Measuring Students' Identification with School", *Educational and Psychological Measurements*, Vol. 56, 760-770.
- Voelkl, K. E. (1997). "Identification with School", *American Journal of Education*, Vol. 105, 294 – 317.
- Warm, T. (1989). "Weighted Likelihood Estimation of Ability in Item Response Theory", *Psychometrika*. Vol. 54, 427-450.

7.0 Survey Methodology

Definition of the YITS population

The YITS target population for the 15 year-old Reading Cohort comprises residents of the ten provinces of Canada who were born in 1984. A large portion of the questionnaire for Cycle 2 that was administered from January to April 2002, is devoted to profiling these individuals' education and labour market activities during the reference year of 2001, when they were 17 years-old. Note that the YITS is strictly a longitudinal survey. The initial sample of 15 year-olds selected at cycle 1 will continue to be surveyed every two years for as many as five cycles. No attempts were made to top-up the sample at cycle 2 to ensure a cross-section representation of the population. Note also that the YITS uses a funnel approach meaning that non-respondents at a specific sample are not followed-up for subsequent cycles of the survey.

At cycle 1, a parent questionnaire was also administered to the parents of the 15 year-olds who participated in the study. There was no such questionnaire at cycle 2. Only the students who participated a cycle 1 were part of the cycle 2 sample.

Please refer to the YITS 15-year-old User Guide, 2000 (Project 5058) at www.statcan.ca for all information on the methodology and sample design used at cycle 1.

The following table shows the sample sizes by province and by cycle. Note that some respondents may have moved to another province or outside of Canada between cycles 1 and 2. The table below shows the sample distribution based on the cycle 1 province or residence. Note also that since our target population is comprised of individuals living in Canada at cycle 1, out-of-country respondents remain in-scope for future cycles. Although they may be more difficult to contact and trace, they were still sent out for cycle 2 collection. There were few such cases.

Table 7.1 15 year-old sample allocation by province

Province	Sample Size at Cycle 1	Sample Size at Cycle 2
Newfoundland and Labrador	2,555	2,281
Prince Edward Island	1,844	1,632
Nova Scotia	3,320	2,930
New Brunswick	3,301	2,963
Quebec	5,024	4,497
Ontario	5,557	4,290
Manitoba	2,955	2,599
Saskatchewan	2,971	2,716
Alberta	3,137	2,742
British Columbia	3,611	3,037
All provinces	34,275	29,687

8.0 Data Quality

8.1 The Frame

The frame for this survey was constructed during the design stage prior to cycle 1 collection. Since the YITS is strictly a longitudinal survey, there were no frame issues specific to cycle 2. For quality issues relating to the frame of 15 year-olds used at cycle 1, please refer to the cycle 1 User Guide.

8.2 Measuring Sampling Error

The estimates derived from this survey are based on a sample of schools and students. 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 15 year-old cohort focused on the desire to obtain reliable estimates of key variables at super-stratum level (see section 9.4 of cycle 1 User Guide). For future cycles, there was also a desire to produce reliable provincial estimates for key characteristics of student leavers and non-leavers. The following results relate to the first of these survey objectives. In order to be able to produce reliable estimates for 5 cycles of these respondents, one would expect CVs for cycle 2 to be well below the maximum allowable targeted CV. For the purpose of analyzing the overall quality of the estimates, the coefficients of variation for a number of key variables were computed for each super-strata and the results are summarized in the Tables 8.1a), 8.1 b) and 8.1 c) and Table 8.3. All characteristics tabulated were for proportions of respondents who fall within a given category. If a question did not apply to the whole population, the proportion was estimated for the subset of the population to which the question applies. The YITS variables used in the analysis are the same ones that were used for the cycle 1 User Guide. This will allow the reader to monitor the situation over the cycles. The cycle 1 variables used were the following:

- A10:** Do you expect to stay in school until you graduate from high school?
A11-YSDV_A11: Derived variable: What is the highest level of education you would like to get?
D2A: Think about your closest friends. How many of these friends ... think completing high school is very important? Response categories: none, some, most, all.
D2D: Think about your closest friends. How many of these friends ... are planning to further their education or training after leaving high school? Response categories: none, some, most, all.
G2A: Did you ever do any work ... for pay for an employer (such as at a store or restaurant)?
G2B: Did you ever do any work ... for pay at an odd job (such as babysitting or mowing a neighbour's lawn)?
G2C: Did you ever do any work ... on your family's farm or in your family's business (with or without pay)?
G11A: Since the beginning of this school year, have you done any work ... for pay for an employer (such as at a store or restaurant)?

- G11B:** Since the beginning of this school year, have you done any work ... for pay at an odd job (such as babysitting or mowing a neighbour's lawn)?
- G11C:** Since the beginning of this school year, have you done any work ... on your family's farm or in your family's business (with or without pay)?
- L2 -YSDV_L2:** Derived variable: What is your approximate overall mark this year?

Tables 8.1a) to c) summarize the CVs computed for all of those estimates. The first table presents the CVs for all the estimates where the estimated proportion falls between 10% and 20% whereas the second table presents results for proportions that fall in the 20% to 30% range. Finally, the third table summarizes all estimates where the estimated proportion is greater than 30%. The results are broken down this way because of the fact that the CV is very dependent on the estimate itself. The lower the estimated proportion, the more likely the CV will be large simply because the denominator in the calculation of the CV is the estimate itself. This phenomenon can be observed in the tables. The CVs in table 8.1a) are on average slightly higher than the ones in table 8.1b), which in turn are slightly higher than the ones in table 8.1c). However, in almost all cases we see that the CV is well below the 16.5% upper limit set at the sample design stage, and that is true for both the cycle 1 and cycle 2 estimates. In fact, very few CVs are higher than 10%. This is expected as the design of the sample aims for CVs below the 16.5% range for five cycles of the survey. Sample attrition will surely continue to affect future cycles of the survey, leading to fewer usable records in estimation and consequently larger sampling errors. Note that the CVs tend to be larger for French students in Nova Scotia and Manitoba. Although the sampling fraction of schools and students for those two super strata was very high, the replication method used to compute the sampling errors assumes the sample was selected with replacement. This will have the effect of over estimating the sampling error for strata where the sampling fraction was high. See Section 12 for more detail on the estimation of sampling variance.

Table 8.1a): Summary of CVs for student variables with $10\% \leq p < 20\%$

Super strata	Cycle 1				Cycle 2			
	Min CV	Max CV	Mean CV	# variables	Min CV	Max CV	Mean CV	# variables
NL_B	4.9	6.8	5.8	6	5.1	7.1	6	6
PE_B	6.2	6.4	6.3	3	6.4	6.7	6.6	3
NS_E	5.5	5.8	5.6	3	5	6.3	5.8	4
NS_F	13.8	18.4	16.1	5	16.6	21.8	18.8	5
NB_E	5.9	6.5	6.2	3	6.5	6.8	6.6	3
NB_F	6.6	8	7.1	4	6.8	9.5	8	5
QC_E	7.1	8.7	8.1	4	7.3	9.3	8.7	4
QC_F	4.7	7.2	6.2	6	4.9	7.3	6.3	6
ON_E	4.4	6.4	5.4	5	4.8	7	5.7	5
ON_F	7.5	10.4	9.3	5	8.1	11.5	10	6
MN_E	5	5.7	5.4	3	5.1	6.1	5.7	3
MN_F	14.2	19	16.6	2	14.4	20.1	17.2	2
SK_B	5	5.4	5.2	2	5.2	6.8	5.8	3
AL_B	5.2	6	5.6	3	4.5	6.3	5.5	4
BC_B	4.8	7.4	5.9	4	4.4	8	5.9	5

Table 8.1b): Summary of CVs for student variables with $20\% \leq p < 30\%$

Super strata	Cycle 1				Cycle 2			
	Min CV	Max CV	Mean CV	# variables	Min CV	Max CV	Mean CV	# variables
NL_B	3.8	4.9	4.2	5	3.9	5	4.3	5
PE_B	3.8	4.9	4.3	6	4	5.2	4.5	6
NS_E	3.9	4.7	4.3	8	4	4.9	4.5	6
NS_F	11.1	12.1	11.7	5	11.9	14.7	13.4	5
NB_E	3.8	5	4.3	8	4.1	5.2	4.5	8
NB_F	4.7	5.3	5	4	5.2	6.5	5.6	4
QC_E	4.7	8.8	6.3	4	4.9	9.4	6.6	4
QC_F	3.7	4.5	4.1	4	3.7	4.6	4.1	4
ON_E	3.1	4.6	3.8	3	3.1	4.8	3.9	3
ON_F	6.3	7.8	7	3	6.6	7.7	7.1	2
MN_E	3.6	5.5	4.5	6	3.7	5.3	4.5	6
MN_F	9.9	11.9	10.9	5	9.9	12.1	11.1	5
SK_B	3.5	4.3	3.9	5	3.7	4.4	4	5
AL_B	3.8	4.7	4.3	5	4.1	4.7	4.4	4
BC_B	3.3	4.5	3.8	6	3.5	4.8	4	5

Table 8.1c): Summary of CVs for student variables with $p \geq 30\%$

Super STR	Cycle 1				Cycle 2			
	Min CV	Max CV	Mean CV	# variables	Min CV	Max CV	Mean CV	# variables
NL_B	0.7	3.9	2.4	13	0.7	4	2.4	13
PE_B	0.7	3.6	2.5	15	0.8	3.8	2.6	15
NS_E	0.6	2.7	2.1	13	0.7	4	2.3	14
NS_F	2	10.2	6.4	14	2.5	12.2	7.4	14
NB_E	0.8	3.1	2.3	13	0.9	3.2	2.5	13
NB_F	0.9	4.4	3.1	15	1.1	5.1	3.4	14
QC_E	1	6.5	3.9	17	1	6.4	4	17
QC_F	1	4.4	2.6	16	1	4.6	2.7	16
ON_E	0.5	3.4	2.5	17	0.6	3.5	2.6	17
ON_F	0.9	5.3	3.9	17	1	5.8	4.1	17
MN_E	0.7	3.8	2.7	15	0.8	3.7	2.7	15
MN_F	1.9	9.5	6.3	15	1.8	9.6	6.4	15
SK_B	0.6	3.4	2.3	16	0.6	3.4	2.4	16
AL_B	0.6	4.8	2.8	17	0.7	4.9	2.9	17
BC_B	0.6	3.6	2.5	15	0.6	3.9	2.6	15

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 4.0**).

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 generally low for the YITS cycle 2. Table 8.2 summarizes partial non-response for the 15 year-old cohort of the cycle 2 questionnaire. In general, item non-response was not a significant problem, although there are a few questions for which the rate was very high. Note that these rates do not include income variables and derived variables that are usually dependent on more than one question and could therefore have slightly higher non-response rates. Also, table 8.2 only presents item non-response rates for variables with less than 50% of values coded as valid skip and for module of the questionnaire where there are at least 5 variables that meet these criteria. The codebook should be used to get non-response rates for specific variables.

Table 8.2 – Partial Non-Response for 15-year-old Cohort – Cycle 2

Module	Number of variables	Min item NR rate*	Max item NR rate*	Mean item NR rate*	Median item NR rate*
B	19	0.39	1.44	0.89	0.84
C	9	0.98	2.72	1.72	1.47
D	12	1.02	4.82	1.74	1.45
F	67	1.63	4.56	2.04	1.95
H	43	0.00	18.67	1.79	0.33
K	25	0.11	2.93	0.48	0.38
M	34	0.80	71.84	23.75	5.18
N	8	2.47	2.54	2.50	2.49
P1	64	0.00	0.38	0.06	0.04
P2	21	0.00	1.60	0.50	0.00
P6	11	1.61	4.10	3.33	3.67
PS	98	0.11	4.10	1.44	0.11
Q	20	3.46	6.47	5.00	5.56
U	27	0.85	9.57	3.76	3.64
V	12	0.10	0.10	0.10	0.10

8.4 Response rates

Table 8.3 presents response rates by province. These rates are not weighted and use, as a base, the initial YITS sample count. Since YITS is longitudinal using a funnel approach, this means that only responding units from cycle 1 were followed in cycle 2. As a result, the initial sample size for cycle 2 was 29,687. 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). Table 8.3 shows provincial response rates for cycle 1, cycle 2, as well as the longitudinal response rate.

The province-level response rates in presented show considerable variation. Ontario shows the lowest overall longitudinal response rate. This is mainly due to a cycle 1 issue where the parental consent was required in order to collect data for the Ontario students. See the cycle 1 User Guide for more detail.

Table 8.3 Provincial Response Rates – Cycles 1 and 2

Province	Response rate (%) Cycle 1	Response rate (%) Cycle 2	Longitudinal Response Rate (%)
Newfoundland	89.7	94.8	85.0
Prince-Edward Island	88.5	91.1	80.6
Nova-Scotia	88.3	89.2	78.8
New-Brunswick	89.8	85.0	76.3
Quebec	89.5	91.8	82.2
Ontario	78.2	90.0	70.4
Manitoba	88.0	93.5	82.3
Saskatchewan	91.4	93.3	85.3
Alberta	87.4	90.8	79.4
British-Columbia	84.1	87.3	73.4
All provinces	86.6	90.5	78.4

9.0 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-neighbour 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 Table 8.1 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 9.1: Imputation rates for income and earnings

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.

Derived Variable	Cohort A		
	Number of imputed values	# records where questions apply	Imputation rate (%)
From module P2:			
Ephei2	2245	48769	4.6
Epwei2	2245	48769	4.6
Epmei2	2245	48769	4.6
Ephsi2	3565	48769	7.3
Epwsi2	3565	48769	7.3
Epmsi2	3565	48769	7.3
Nwwpmei2	627	50755	1.2
Nhwpmei2	1089	50755	2.1
Nwwpmsi2	914	50755	1.8
Nhwpmsi2	1660	50755	3.3
From module V:			
Incwsd2	0	26880	0.0
Incsed2	0	26880	0.0
Incsgd2	1045	26880	3.9
Inceid2	1150	26880	4.3
Incstd2	1315	26880	4.9
Incspd2	1042	26880	3.9
Incsd2	1007	26880	3.7
Incnrdd	2958	26880	11.0
Inctbd2	1489	26880	5.5
Incgd2	1040	26880	3.9
Incond2	1233	26880	4.6
Tincd2	3669	26880	13.6
Incwssd2	68	239	28.5
Incsesd2	23	239	9.6
Incsgsd2	12	239	5.0
Inceisd2	26	239	10.9
Incstd2	40	239	16.7
Incspsd2	14	239	5.9
Incsd2	7	239	2.9
Incnrdsd2	30	239	12.6
Inctbsd2	11	239	4.6
Incgssd2	16	239	6.7
Inconsd2	14	239	5.9
Tincsd2	97	239	40.6

10.0 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.

10.1 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.

10.2 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.

10.2.1 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

10.2.2 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.

10.2.3 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 Section 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).

10.3 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 Section 11.0.

10.4 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 Section 8.0. However, for this purpose, the quality level of an estimate will be determined only based on sampling error as reflected by the coefficient of variation as shown in the table below. Nonetheless, users should be sure to read Sections 8.0 through 8.4 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.

Table 10.1 Quality Level Guidelines

Quality Level of Estimate	Guidelines
1. Acceptable	<p>Estimates have: a sample size of 30 or more, and coefficients of variation in the range 0.0% - 16.5%</p> <p>No warning is required.</p>
2. Marginal	<p>Estimates have: a sample size of 30 or more, and coefficients of variation in the range 16.6% - 33.3%.</p> <p>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.</p>
3. Unacceptable	<p>Estimates have: a sample size of less than 30, or coefficients of variation in excess of 33.3%.</p> <p>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.</p>

11.0 Weighting

11.1 Cycle 2 Student Weight

The starting point for creating the cycle 2 weights was the final cycle 1 weights for respondents of the 15 year-old cohort. The details of how the cycle 1 weights were derived are provided in the cycle 1 User Guide. The purpose of the cycle 2 weight adjustment is strictly to account for non-response that occurred during cycle 2. To account for people who participated in cycle 1 and did not participate in cycle 2, the final cycle 1 weights of those who participated in both cycles were proportionally increased so that the sum of their adjusted weights would equal the sum of the cycle 1 final weights. Although response rate for cycle 2 was high, analysis of non-response patterns showed that the non-response adjustments should take into consideration certain variables. The adjustments were made separately within response classes defined by those variables. Variables used included province, language, reading scale level at cycle 1, overall marks at cycle 1, family structure and a dichotomous variable identifying acceptance by the parents to share their data at cycle 1.

Note also that respondents deceased in cycle 2 were treated as respondents since they represented others in the target population that have died since cycle 1 but were not in the sample. For these deceased respondents, their cycle 2 data was set to missing and a flag called DECEASE2 was created in order to easily identify these cases. The deceased do have a longitudinal weight and the reason they are kept on the file is so that the sum of the weights is consistent with the sum of the weights from cycle 1. These records should be removed from most analyses since they do not contain any information from the cycle 2 questionnaire.

In addition to the final student weight, some additional weights were created for cycle 2. These were required for a subset of the sample and are explained below. Note that full sample student weights are often referred to as the reading weights. The reason for that is that reading was the main domain of interest for PISA 2000. Every student selected in the sample was tested for reading. The Mathematics and Science sub-domains were not tested on every student. See Section 10.3 for more detail on the sub-domains.

11.2 Cycle 2 Parent Weight

Although there was no parent questionnaire in cycle 2, a parent weight was derived to allow analysis of cycle 2 student characteristics combined with parent data collected at cycle 1. A subset of the cycle 2 student characteristics containing only those records where the parents had filled out a questionnaire at cycle 1 was created. This subset contains 24,397 records. This set of data was weighted using the same approach used for the full sample weights. In fact, because the student weighting approach used the parent-sharing variable from cycle 1 in the non-response adjustment model, the same model used was essentially used for weighting the parent data at cycle 2.

11.3 Sub-Domain Weights

When YITS and PISA were jointly administered in 2000, every student was tested for reading ability since reading was the major domain for PISA 2000. For the mathematics and science sub-domains, only 5/9 of the sample students were tested in each of these domains. Operationally, there were 9 PISA exam booklets in total and all 9 covered Reading whereas 5 of the 9 covered Mathematics and 5 of the 9 covered Science. This meant that 1 of the 9 exam booklets covered all three domains of study. Consequently, a Mathematics and a Science weight had to be derived at cycle 1 to allow for inference for those sub-samples. Although there were no such sub-groups for cycle 2 (all respondents were asked all questions), there may still be interest in cross tabulating some of the sub domain data obtained at cycle 1 with some of the cycle 2 data. Similarly, there could be interest in looking at cycle 2 characteristics with parent data for a sub-domain. Therefore, in addition to the cycle 2 student weights (section 10.1) and cycle 2 parent weight (Section 10.2), four other sets of weights were also produced. Table 11.1 summarizes the various weights for cycle 2 – 15 year-olds, Cohort A.

Table 11.1 Cycle 2 Weights – 15 year-olds

File Name	File description	# records	Weight	Replicate Weights
cycle2_std_read_wgts	Full sample weights. This file contains all respondents who participated in cycle 2 as well as deceased respondents.	26,880	w2_ysr	b2sr1 to b2sr1000
cycle2_std_math_wgts	Math sub-sample weights. Contains respondent who participated in cycle 2 and were part of the math sub-sample at cycle 1.	14,980	w2_ysm	b2sm1 to b2sm1000
cycle2_std_science_wgts	Science sub-sample weights. Contains respondent who participated in cycle 2 and were part of the science sub-sample at cycle 1.	14,920	w2_ysc	b2ss1 to b2ss1000
cycle2_par_read_wgts	Parent full sample weights. Contains respondents who participated at cycle 2 and who's parents participated at cycle 1.	24,397	w2_ypr	b2pr1 to b2pr1000
cycle2_par_math_wgts	Parent math sub-sample weights. Contains respondents who participated at cycle 2, who were part of the math sub-sample at cycle 1 and who's parents participated at cycle 1.	13,606	w2_ypm	b2pm1 to b2pm1000
cycle2_par_science_wgts	Parent science sub-sample weights. Contains respondents who participated at cycle 2, who were part of the science sub-sample at cycle 1 and who's parents participated at cycle 1.	13,551	w2_yps	b2ps1 to b2ps1000

12.0 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 15 year-old Reading 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 Section 11.0). 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 15 year-old Reading Cohort a total of 1,000 replicates were created at cycle 1 and those same replicate weights are the starting point for deriving the cycle 2 replicate weights. Each replicate was initially 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. The cycle 2 final Bootstrap replicate weights are summarized in table 11.1 of section 11.3 on weighting.

13.0 Working with YITS files

13.1 Roster and Flat Files

File or Roster	Cohort A – 15 year-olds (17 years-old in 2001)	Cohort B – 18-20 year-olds (20-22 years-old in 2001)
Main Flat File	Cycle 2 - Cohort A - Person Level Main File	Cycle 2 - Cohort B - Person Level Main File
K roster	Cycle 2 - Cohort A - Post Secondary Engagement Roster	Cycle 2 - Cohort B - Post Secondary Engagement Roster
Hinst roster	Cycle 2 - Cohort A - Education above High School (Institution Roster)	Cycle 2 - Cohort B - Education above High School (Institution Roster)
Hprog roster	Cycle 2 - Cohort A - Education above High School (Program Roster)	Cycle 2 - Cohort B - Education above High School (Program Roster)
P1cycle1 roster	N/A	Cycle 2 - Cohort B - Confirmation of Open Jobs from Cycle 1 Roster
P1cycle2 roster	Cycle 2 - Cohort A - Job Roster	Cycle 2 - Cohort B - Job Roster
P2 roster	Cycle 2 - Cohort A - Job Details Roster	Cycle 2 - Cohort B - Job Details Roster

13.2 Youth In Transition Survey: Data Extraction Tool

13.2.1 About Youth in Transition Survey (YITS)

The Youth in Transition Survey (YITS) is a longitudinal survey undertaken jointly by Statistics Canada and Human Resources and Skills Development Canada. This survey is designed to examine the major transitions in the lives of youth, particularly between education, training and work.

The YITS is designed to examine the patterns of, and influences on, major transitions in young peoples' lives, particularly with respect to education, training and work. Human Resources and Skills Development Canada and Statistics Canada have been developing the YITS in consultation with provincial and territorial ministries and departments of labour and education. Content includes measurement of major transitions in young people's lives including virtually all formal educational experiences and most about-market experiences, achievement, aspirations and expectations, and employment experiences. The implementation plan encompasses a longitudinal survey of each of two cohorts, ages 15 and 18-20, to be surveyed every two years.

The results from the Youth in Transition Survey will have many uses. Human Resources and Skills Development Canada will use them to aid policy and program development. Other users of the results include educators, social and policy analysts, and advocacy groups. The information will show how young adults are making their critical transitions into their adult years.

These researchers and analysts will have access to important information that can be used in developing programs to deal with both short-term and long-term problems or barriers that young adults may face in their pursuit of higher education or in gaining work experience. Information from the survey will help to evaluate the effectiveness of existing

programs and practices, to determine the most appropriate age at which to introduce programs, and to better target programs to those most in need.

Young adults themselves will be able to see the impact of decisions relating to education or work experiences. They will be able to see how their own experiences compare to those of other young adults.

13.2.2 Statistical Activity

PISA/YITS is one project, which consists of two parallel survey programs: the Programme for International Student Assessment (PISA) and the Youth in Transition Survey (YITS).

PISA is an international assessment of the skills and knowledge of 15 year-olds which aims to assess whether students approaching the end of compulsory education have acquired the knowledge and skills that are essential for full participation in society.

YITS is designed to examine the patterns of, and influences on, major transitions in young people's lives, particularly with respect to education, training and work. Human Resources and Skills Development Canada and Statistics Canada have been developing the YITS in consultation with provincial and territorial ministries and departments of labour and education. Content includes measurement of major transitions in young people's lives including virtually all formal educational experiences and most about-market experiences, achievement, aspirations and expectations, and employment experiences. The implementation plan encompasses a longitudinal survey of each of two groups, ages 15 and 18-20, to be surveyed every two years.

The 15 year-old respondents to the Reading Cohort (conducted in 2000) participated in both PISA (Survey 5060) and YITS (Survey 5058). Starting in 2002, they will be followed up longitudinally by YITS (Survey 4435).

The 15 year-old respondents to the Mathematics Cohort (conducted in 2003) participated in both PISA (Survey 5060) and YITS (Survey 5059). They will not be followed up longitudinally.

13.2.3 Purpose of the Application

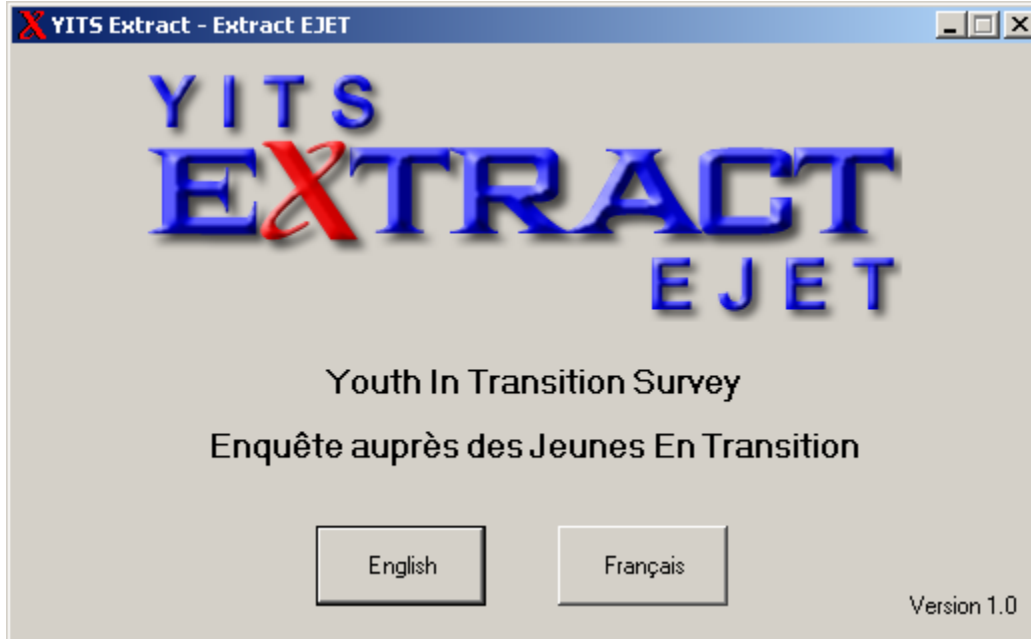
The YITS data sets are many, large and are stored two different formats (normalized data for the main file with rostered or un-normalized data for information collected that have many iterations). As the survey continues and more cycles of information are available the number of files and complexity in figuring out how to use the files will grow to a point where it will take more time programming and running a merge sequence than actually researching.

The YITS: Data Extraction Tool will facilitate the process users go through to create their files used in research and analysis. The application provides an intuitive and direct interface for users to select the specific variables they need in order to produce their findings. Behind the interface the program will normalize the information that is in roster format, merge each of the individual data files from which variables were selected and assign the weight file based on the cycle and cohort choices of the user. Due to the complexity of assigning weights (i.e. depending on what variables and cohort are selected there is a lot of room for error) the program will also automatically assign the appropriate bootstrap weights to the file. The design of the application also takes into account the longitudinal aspect of the survey when merging the data files, the number of observations in the final data file will be based on the most recent cohort selected by the user.

This application will do much to facilitate the initial use of YITS data, unfortunately it isn't a tool to explain how to use the survey in analysis or research. The data files are merged according to the design of the survey; to understand why the tool was necessary along

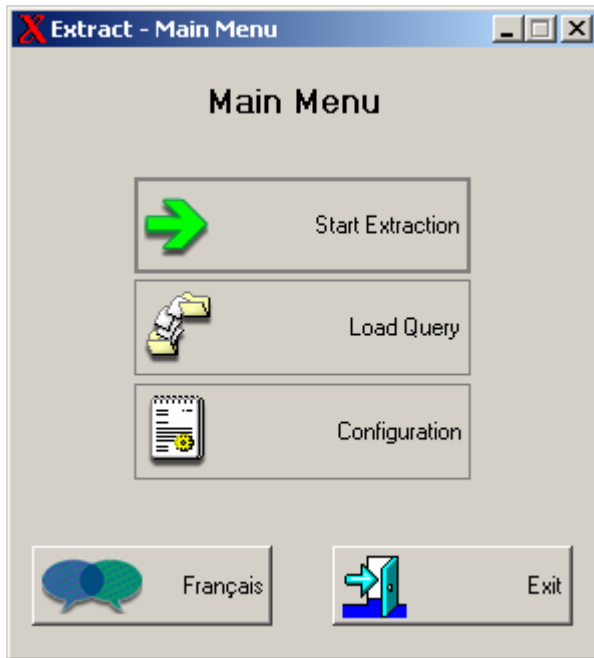
with the overall design of PISA & YITS the researcher must consult other materials. Using the data extraction tool in conjunction with the codebooks, questionnaires and user guides will provide a researcher with a solid foundation for their work. Other reference materials that may be used are available on the Statistics Canada website under the link for definitions, data sources and methods section (<http://www.statcan.ca/english/sdds/0020t.htm>).

Splash Page



Welcome to the Extraction Tool: The language selection buttons serve two purposes: selecting the language of the application and determining the language of the output file and formats

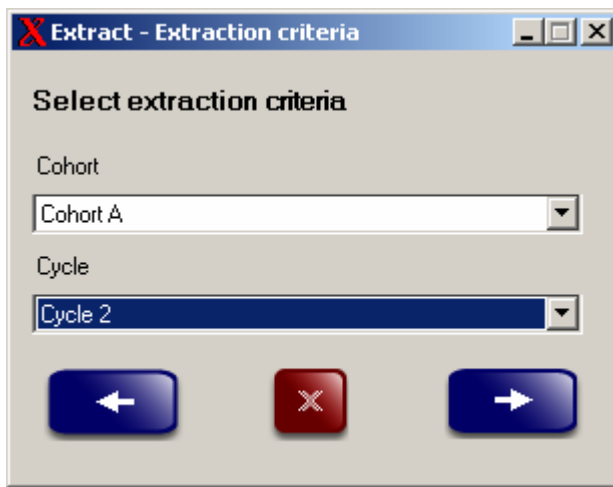
Main Menu



Main Menu: This is the central menu for the application. From this point you can start a new extraction, load a previously saved extraction routine, change the language of the extraction, close down the application and go into the configuration settings. Please note that the users query can be saved in the “Select variables menu”. Also, a query that had previously been saved in a particular language will work even if the language of the application has changed. When the user cancels their extraction in subsequent menus the application automatically brings them back to the main menu. The configuration settings are for the local administrator to use

Users must be linked to the YITS Data Extraction Tool using the installation procedures found in the YITS Administration Documentation.

Extraction Criteria



Primary PISA/YITS Extraction Criteria: The first of two Extraction Criteria Menus where users can specify their population of interest.

Cohorts:

- Cohort A – 15 years old in year 2000
- Cohort B – 18-20 years old in year 2000

Cycle:

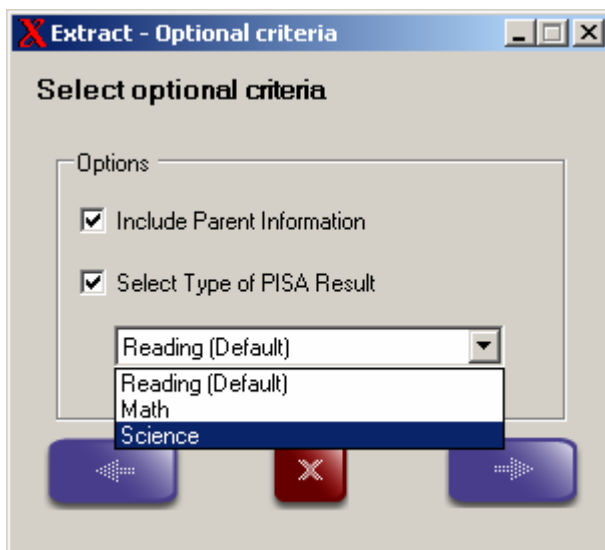
- Cycle 1 – Survey year 2000
- Cycle 2 – Survey year 2002
- Cycle 3 – Survey year 2004

Navigation Buttons:

- Back arrow – Main menu
- X – Main menu
- Forward arrow:
 - If cohort A selected – Optional Criteria menu
 - If cohort B selected – Select Data menu

For more information on the cohorts and cycles please refer to the user guide or under Definitions, Data Sources and Methods on the Statistics Canada website.

Optional Extraction Criteria



Optional Extraction Criteria Menu: If Cohort A(15 year-olds) is selected from the Primary Extraction Criteria menu the Optional Criteria menu opens.

Users can:

- Select the parent information
- Choose which PISA test results they wish to use

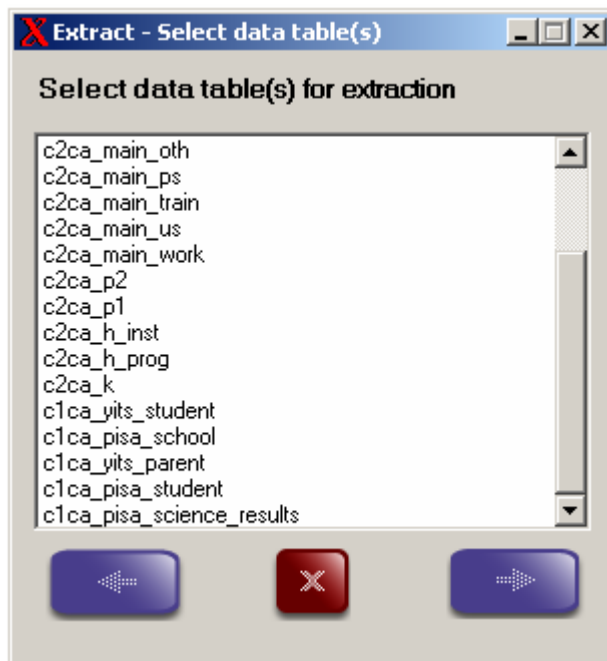
Navigation Buttons:

- Back arrow – Extraction Criteria menu
- X – Main menu
- Forward arrow - Select Variables menu

For each cycle of PISA/YITS there are seven sample weights to choose from, each with 1000 bootstrap weights for calculating the variance estimates. The Extraction Criteria menus allow the user to select any of the possible paths a respondent could have followed and then assigns the appropriate weight to the output data file based on what the user has chosen to look at.

It is important to note that if a user does not wish to have parent information or test results included in their analysis file they do not have to select anything. The reading weights are assigned to the file by default (according to the design of the survey). Selecting parent information or either the Mathematics or Science results will have a significant impact on the sample size of the output data file. Each of the options presented in the Optional Extraction Criteria Menu are a sub sample of the PISA/YITS population. For more information on how each of the selections may affect the sample extracted please refer to Section 8.0 Data Quality in this User Guide or under Definitions, Data Sources and Methods on the Statistics Canada website under Record Numbers 4435, 5058 or 5059.

Select Data Tables



Selecting Data Files: Once the user has completed the Extraction Criteria menus the application presents the user with a list of data files to choose from. If the user chose to use information for a later cycle of the survey, all data files from previous cycles of PISA/YTIS will be made available for browsing.

For the Record Number 4435 (YITS), The Main Person Level file has been broken down into multiple data sets of about 100 variables each for ease of use with this application. This has been done for Cohort B (18 to 20 year-olds) in Cycle 1 and for both Cohorts (17 and 20 to 22 year-olds) in Cycle 2.. Each of the data files have been loosely grouped around the modules of the YITS questionnaire. For example, the data file “c2ca_main_us” contains variables related to the questions about moving to the US – Module A.

For a complete list of the data files and the variables contained within refer to the Statistics Canada website under Definitions, Data Sources and Methods for Record Numbers 4435, 5058 or 5059.

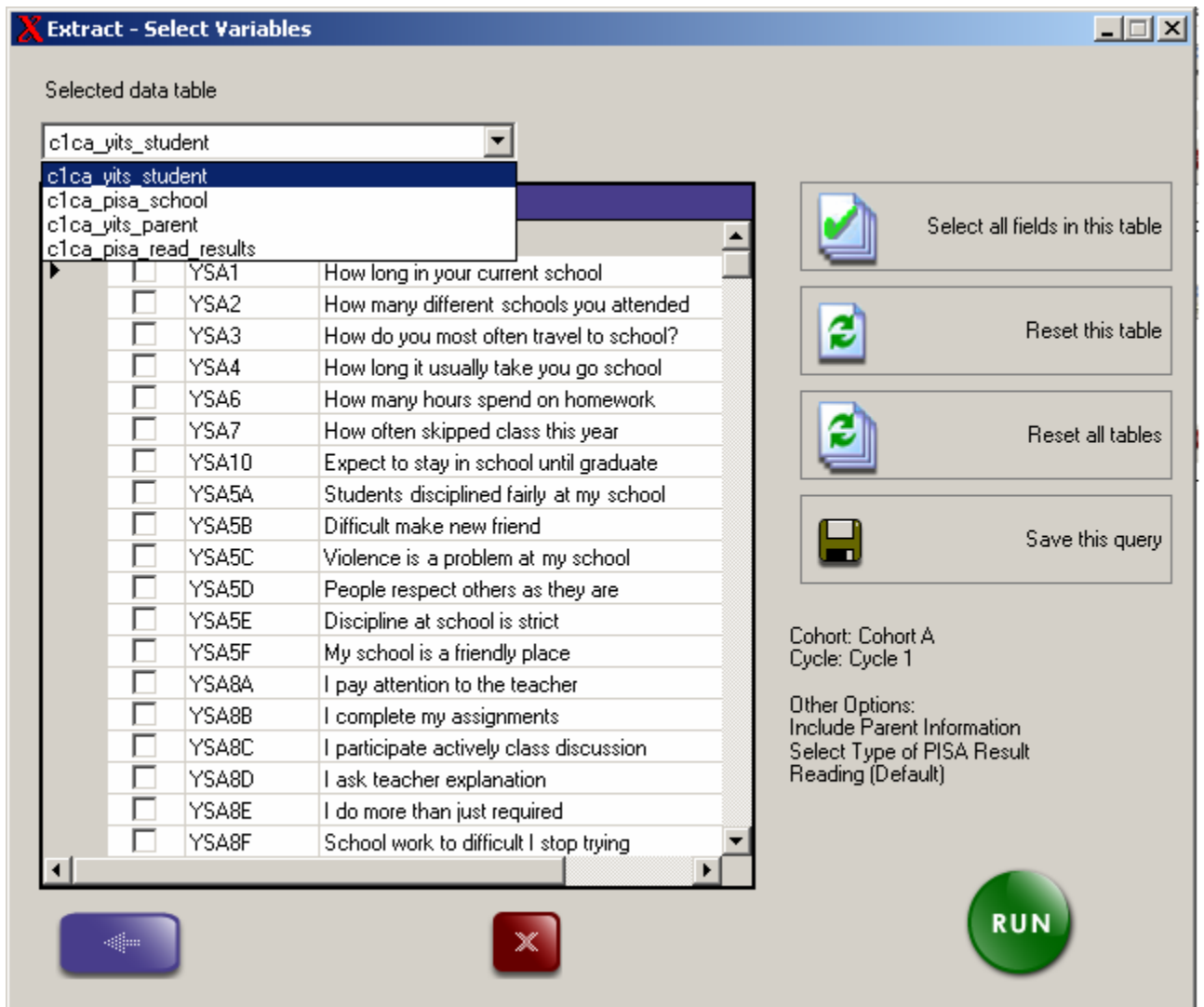
Naming convention for the data files:

- **c2ca** - Cycle 2 Cohort A;
- **c1cb** – Cycle 1 Cohort B
- **_Main** - indicates that data file is part of the main respondent information (Note if _Main isn't included in the cycle 2 file name it is a rostered file)
- **_dem, _fund, _work** etc. – indicates which subject or module is represented in the particular data file
- **_yits** or **_pisa** – are for cohort A indicating whether the information is from the PISA or YITS component of the survey.

Navigation Buttons:

- Back arrow
 - If cohort A selected – Optional Criteria menu
 - If cohort B selected – Extraction Criteria menu
- X – Main menu
- Forward arrow - Select Variables menu

Select Variables



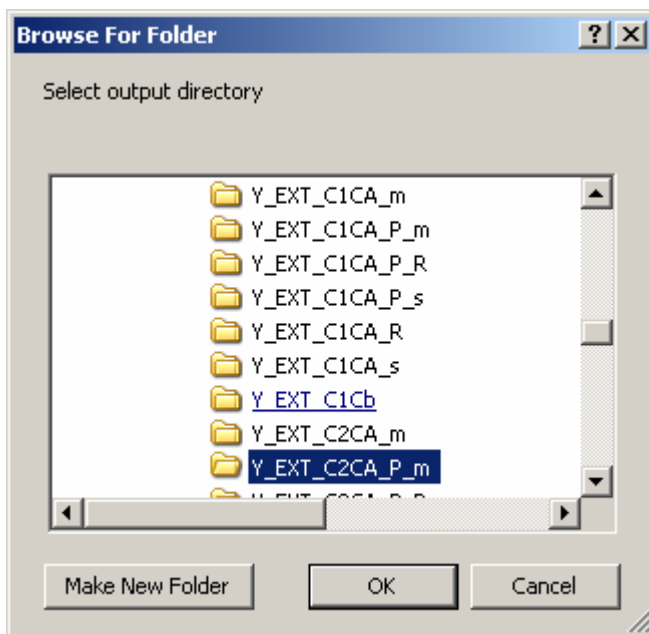
Selecting Variables for output:

- Drop down menus allows user to select the file from which they can choose their variables.
 - There is a counter next to the data file name to show how many variables have been selected.
- Some buttons have been provided to aid the user in:
 - Selecting all variables in the selected file
 - Resetting their choices for the selected file
 - Resetting the entire variable selection process.
- Save Query button allows the user to save their current extraction set-up so that they can update their extraction when further information is necessary.
- Notes are provided for the user to document from which cycle and cohort data are being extracted.

Navigation Buttons:

- Back arrow – Select Data menu
- X – Main menu
- Run – Starts the extraction process, opens the Browse for output folder. Please Note: Pressing the RUN button without selecting any variable from any table creates a file containing only the weights and the bootstrap weights.

Output Folder



Selecting the output directory:

- Allows users to output the information to a specific directory
- If there are data files or other information already in the output directory there will be a warning that the program may overwrite a data file.
- The output is in the form of
 - SAS & SPSS cards, set to run in the output folder selected.
 - An ASCII file (.dat) that contains the variables selected and along with the appropriate weights (the sample weight and 1000 bootstrap weights).

13.2.4 Saving and Loading Queries

Using the output files

The extraction tool produces four output files for each run:

- One data file in ascii format containing the variables of interest and the weight variables. The name of the data file is generated automatically by the program
- Two SAS programs are generated:
 - the program with the prefix “Create” will generate the SAS data file with all of the available formats & labels applied to it.
 - The SAS program with the prefix “Lrecl” contains the record layout for the ascii file and is referenced by the first SAS program.
- The SPSS program also generated by the extraction tool will create an SPSS file from the ascii data file.

Example of output

Ascii file containing information from: Cycle 2 Cohort A (C2CA), selected parent variables (_PAR), selected PISA variables (_PISA), English labels and formats (_E), with Math results (_F)

- C2CA_PAR_PISA_E_M.DAT

The Sas files :

- Create_C2CA_PAR_PISA_E_M.sas
- Lrecl_C2CA_PAR_PISA_E_M.sas

The SPSS Cards

- Create_C2CA_PAR_PISA_E_M.sps

APPENDICES

APPENDIX A – Cycle 2 – New “Other – Specify” Categories

Module and Variable	New Category or New question from “other-specify”	English Description
Module D: D2Q05	11 12	Other - To gain experience/knowledge Other - Money
D2Q06	10 11 12 13 14	Other - academic workload heavy/more important Other - already participated Other - conflict with courses/schedule Other - no time Other - transportation
Module F : (Mark all that apply categories – new from F2S63F) F2Q63FI F2Q63FJ F2Q63FK	I J K	Other - Did not need work. Other - Could not get to work/location Other - Foreign student (no visa/work permit)”
Module H : H2Q411	09	Other - Offered only later/higher grades
Module L : (Mark all that apply categories - new from L2S08) L2Q03A	L2Q03A	Other...had you used money from your personal savings? (exclude money earned from jobs since leaving high school).
Module M : M2Q02	10 11 12	Other - Will apply Other - Future undecided Other - Need information on program
Module P5 : P52Q06	12 13 14 15	Other – Satisfied Other - House/lease Other - Don't want/need to Other - Age
Module P6 : (Mark all that apply categories –new from P62S25) P62Q25K	K	Other - Age

APPENDIX B – Cycle 1 – Module H variables

Cycle 1, module H variables

Person level variables

HGDA	HGDAA	HLPS	HEDAT	HEDL	DLPSM	DLPSY	DLPSFM	DLPSFY
NINDI	NPRDI	LPSAT	MHSPS	MHSPSFLG	AGSPS	FPSP	EDTPSM	EDTPSY

Institution level variables

Questionnaire variables

Institution 1	Institution 2	Institution 3	Institution 4
H8a	H8b	H8c	H8d
H9a	H9b	H9c	H9d
H10Aa_1	H10Ab_1	H10Ac_1	H10Ad_1
H10Aa_2	H10Ab_2	H10Ac_2	H10Ad_2
H12a	H12b	H12c	H12d

Derived variables

NPRPI1	NPRPI2	NPRPI3	NPRPI4
DSAINM_1	DSAINM_2	DSAINM_3	DSAINM_4
DSAINY_1	DSAINY_2	DSAINY_3	DSAINY_4
DLINM_1	DLINM_2	DLINM_3	DLINM_4
DLINY_1	DLINY_2	DLINY_3	DLINY_4
FPLIN_1	FPLIN_2	FPLIN_3	FPLIN_4
DLFINM_1	DLFINM_2	DLFINM_3	DLFINM_4
DLFINY_1	DLFINY_2	DLFINY_3	DLFINY_4
HLATT_1	HLATT_2	HLATT_3	HLATT_4

Program level variables

Questionnaire variables

Institution 1			Institution 2			Institution 3		Institution 4
program 1	program 2	program 3	program 1	program 2	program 3	program 1	program 2	program 1
H18a1_B	H18a2_B	H18a3_B	H18b1_B	H18b2_B	H18b3_B	H18c1_B	H18c2_B	H18d1_B
H18a1_C	H18a2_C	H18a3_C	H18b1_C	H18b2_C	H18b3_C	H18c1_C	H18c2_C	H18d1_C
H21a1	H21a2	H21a3	H21b1	H21b2	H21b3	H21c1	H21c2	H21d1
H22a1	H22a2	H22a3	H22b1	H22b2	H22b3	H22c1	H22c2	H22d1
H23a1	H23a2	H23a3	H23b1	H23b2	H23b3	H23c1	H23c2	H23d1
H26Aa1	H26Aa2	H26Aa3	H26Ab1	H26Ab2	H26Ab3	H26Ac1	H26Ac2	H26Ad1
H26Ba1	H26Ba2	H26Ba3	H26Bb1	H26Bb2	H26Bb3	H26Bc1	H26Bc2	H26Bd1
H29a1	H29a2	H29a3	H29b1	H29b2	H29b3	H29c1	H29c2	H29d1
H30_1Mth	H30a2_Mt	H30a3_Mt	H30b1_Mt	H30b2_Mt	H30b3_Mt	H30c1_Mt	H30c2_Mt	H30d1_Mt
H30_1Yr	H30a2_Yr	H30a3_Yr	H30b1_Yr	H30b2_Yr	H30b3_Yr	H30c1_Yr	H30c2_Yr	H30d1_Yr
H39a1	H39a2	H39a3	H39b1	H39b2	H39b3	H39c1	H39c2	H39d1
H42a1	H42a2	H42a3	H42b1	H42b2	H42b3	H42c1	H42c2	H42d1
H43Aa1	H43Aa2	H43Aa3	H43Ab1	H43Ab2	H43Ab3	H43Ac1	H43Ac2	H43Ad1
H43Ba1	H43Ba2	H43Ba3	H43Bb1	H43Bb2	H43Bb3	H43Bc1	H43Bc2	H43Bd1
H44a1	H44a2	H44a3	H44b1	H44b2	H44b3	H44c1	H44c2	H44d1
H45a1	H45a2	H45a3	H45b1	H45b2	H45b3	H45c1	H45c2	H45d1
H48Aa1	H48Aa2	H48Aa3	H48Ab1	H48Ab2	H48Ab3	H48Ac1	H48Ac2	H48Ad1
H48Ba1_1	H48Ba2_1	H48Ba3_1	H48Bb1_1	H48Bb2_1	H48Bb3_1	H48Bc1_1	H48Bc2_1	H48Bd1_1
H48Ba1_2	H48Ba2_2	H48Ba3_2	H48Bb1_2	H48Bb2_2	H48Bb3_2	H48Bc1_2	H48Bc2_2	H48Bd1_2
H48Ba1_3	H48Ba2_3	H48Ba3_3	H48Bb1_3	H48Bb2_3	H48Bb3_3	H48Bc1_3	H48Bc2_3	H48Bd1_3
H48Ba1_4	H48Ba2_4	H48Ba3_4	H48Bb1_4	H48Bb2_4	H48Bb3_4	H48Bc1_4	H48Bc2_4	H48Bd1_4
H48Ba1_5	H48Ba2_5	H48Ba3_5	H48Bb1_5	H48Bb2_5	H48Bb3_5	H48Bc1_5	H48Bc2_5	H48Bd1_5
H48Ba1_6	H48Ba2_6	H48Ba3_6	H48Bb1_6	H48Bb2_6	H48Bb3_6	H48Bc1_6	H48Bc2_6	H48Bd1_6
H49Aa1	H49Aa2	H49Aa3	H49Ab1	H49Ab2	H49Ab3	H49Ac1	H49Ac2	H49Ad1
H49Ba1	H49Ba2	H49Ba3	H49Bb1	H49Bb2	H49Bb3	H49Bc1	H49Bc2	H49Bd1

Derived variables

LVPR_11	LVPR_12	LVPR_13	LVPR_21	LVPR_22	LVPR_23	LVPR_31	LVPR_32	LVPR_41
CLGPR_11	CLGPR_12	CLGPR_13	CLGPR_21	CLGPR_22	CLGPR_23	CLGPR_31	CLGPR_32	CLGPR_41
DLPRM_11	DLPRM_12	DLPRM_13	DLPRM_21	DLPRM_22	DLPRM_23	DLPRM_31	DLPRM_32	DLPRM_41
DLPRY_11	DLPRY_12	DLPRY_13	DLPRY_21	DLPRY_22	DLPRY_23	DLPRY_31	DLPRY_32	DLPRY_41
FPLPR_11	FPLPR_12	FPLPR_13	FPLPR_21	FPLPR_22	FPLPR_23	FPLPR_31	FPLPR_32	FPLPR_41
DLFPRM11	DLFPRM12	DLFPRM13	DLFPRM21	DLFPRM22	DLFPRM23	DLFPRM31	DLFPRM32	DLFPRM41
DLFPRY11	DLFPRY12	DLFPRY13	DLFPRY21	DLFPRY22	DLFPRY23	DLFPRY31	DLFPRY32	DLFPRY41
SIPR_11	SIPR_12	SIPR_13	SIPR_21	SIPR_22	SIPR_23	SIPR_31	SIPR_32	SIPR_41
RSIPR_11	RSIPR_12	RSIPR_13	RSIPR_21	RSIPR_22	RSIPR_23	RSIPR_31	RSIPR_32	RSIPR_41
NMDUR_11	NMDUR_12	NMDUR_13	NMDUR_21	NMDUR_22	NMDUR_23	NMDUR_31	NMDUR_32	NMDUR_41
DSPRM_11	DSPRM_12	DSPRM_13	DSPRM_21	DSPRM_22	DSPRM_23	DSPRM_31	DSPRM_32	DSPRM_41
DSPRY_11	DSPRY_12	DSPRY_13	DSPRY_21	DSPRY_22	DSPRY_23	DSPRY_31	DSPRY_32	DSPRY_41
AGEPS_11	AGEPS_12	AGEPS_13	AGEPS_21	AGEPS_22	AGEPS_23	AGEPS_31	AGEPS_32	AGEPS_41
OPSP_1	OPSP_2	OPSP_3	OPSP_4	OPSP_5	OPSP_6	OPSP_7	OPSP_8	OPSP_9
I1p1MFS1	I1p2MFS1	I1p3MFS1	I2p1MFS1	I2p2MFS1	I2p3MFS1	I3p1MFS1	I3p2MFS1	I4p1MFS1
I1p1MFS2	I1p2MFS2	I1p3MFS2	I2p1MFS2	I2p2MFS2		I3p1MFS2		

LINKS TO REFERENCE DOCUMENTS

The Statistics Canada website is:

<http://www.statcan.ca/english/concepts/index.htm>

Instructions to access survey documentation:

Access the Survey "*Definitions, Data Sources and Methods*" under "*Survey Information*", either alphabetically or by subject:

Alphabetically

- Youth in Transition Survey (YITS) – Project Codes 4435, 5058 and 5059 for Cycles 1 and 2

Subject

- Education (click on "surveys" and the list is alphabetical)

The project codes for YITS are in brackets.

Documentation Available on the website www.statcan.ca:

Questionnaires:

YITS – 18-20 year-olds and 20-22 year-olds (Cycles 1 and 2) (4435)

YITS – 15 year-old Reading Cohort (5058)

- 15 year-old Reading Cohort Questionnaire (Canadian Longitudinal Youth in Transition Survey)
- Parent Questionnaire (Canadian Longitudinal Youth in Transition Survey)

YITS – 15 year-old Mathematics Cohort (5059)

- 15 year-old Mathematics Cohort Questionnaire (Canadian Longitudinal Youth in Transition Survey)
- Parent Questionnaire (Canadian Longitudinal Youth in Transition Survey)

Codebooks:

YITS (4435) – Cycles 1 and 2

YITS (5058) – Cycle 1

- Student Codebook
- Parent Codebook

YITS (5059) – Cycle 2

YITS Data Extraction Tools:

YITS Data Extraction Excel spreadsheets – 4435, 5058, 5059 (provide all file and roster names and variable names for both Cycles 1 and 2)

OTHER DOCUMENTATION AVAILABLE ON REQUEST :

YITS Project Overview (5058 and 4435) – Cycles 1 and 2

The Survey/Project Overview is presented as a mapping document with subject matter themes and also provides the comparison of questions/variables between each cohort of YITS. This document is updated for each occurring cycle of YITS.

Coding Lists for Institutions

ESIS Codes for postsecondary institutions for all provinces and territories.

YITS Data Extraction Tool

Administrative Documentation (to be used for installation of the YITS Data Extraction Tool)