



Microdata User Guide

**National Longitudinal Survey
of Children and Youth**

Secondary File

Cycle 4

September 2000 to June 2001



Statistics
Canada

Statistique
Canada

Canada

Table of Contents

1.0	Introduction	5
2.0	Background	6
3.0	Objectives	8
4.0	Concepts and Definitions	9
4.1	Definitions	9
4.2	Family Derived variables	10
4.3	Respondent	10
5.0	Survey Methodology - Sample	11
6.0	Data Collection	12
6.1	Household collection	12
6.1.1	Child component	12
6.2	School collection	13
6.2.1	Teacher's questionnaire	13
6.2.2	Principal's questionnaire	13
7.0	Data Processing	14
7.1	Editing	14
7.2	Data Capture	14
7.3	CleanUp	14
7.4	PreEdit	14
7.5	Flow Edits	15
7.6	Coding	15
7.7	Consistency Editing	15
7.8	Naming Convention and Coding Structure for NLSCY Variables	15
7.9	Acronym Names for Questionnaire Sections	17
7.10	Examples of Variables Names	19
7.11	Coding Structure for NLSCY Variables	19
7.12	Imputation flags	20
7.13	Derived Variables	21
8.0	Content of the Survey	22
8.1	Processing System	22
8.2	Demographic Variables	22
8.3	Adult Questionnaire	22
8.4	Child Questionnaire	23
8.5	Cautionary Note	25
9.0	Validation of the survey scales	26
9.1	Validation of scale data	26
9.1.1	Scale Definition	26
9.1.2	Scales and Calculations	26
9.1.3	Evaluation of Scale Data	26
9.2	Factor analyses	27
9.2.1	Factor Analysis for Scales	27
9.2.2	Data Transformation Using Optimal Scaling	28
9.2.3	Factor Analysis Using Weighted Data	28
9.3	Calculation of scores and item imputation	28

9.3.1	Calculation of Scores for Each Factor	28
9.3.2	Example of Factor Score Computation	28
9.3.3	Negative Loading	29
9.3.4	Non-Response Code	29
9.3.5	Raw Items.....	29
9.4	Reliability measures for scales	30
9.4.1	Cronbach's Alpha	30
9.4.2	Interpretations of Cronbach's Alpha	30
9.4.3	What is a satisfactory level of reliability?	30
9.5	Cycle 4 Education Scales.....	31
10.0	Survey methodology - Response rates and Coverage of the Education Component	33
10.1	Survey undercoverage	33
10.2	Total Non-response	35
10.2.1	Definitions	35
10.3	Consents	36
10.4	Non-response by teachers and school principals	36
11.0	Imputation	38
11.1	Imputation of responses from school principals	38
11.2	Imputation of the data from the teachers	39
12.0	Weighting and treatment of coverage and non-response	40
12.1	Adjustment of the under-coverage	40
12.2	Adjustment of the parents' consent	41
12.3	Adjustment of the consent of the school boards	42
12.4	Adjustment for non-response by teachers or principals	42
12.5	Summary of adjustments	44
12.6	Limitations	44
13.0	Data quality and Coverage of the Main Survey.....	45
14.0	Guidelines for Tabulation, Analysis and Release	46
14.1	Rounding Guidelines	46
14.2	Sample Weighting Guidelines for Tabulation	46
14.3	Definitions of Types of Estimates: Categorical vs. Quantitative	47
14.3.1	Categorical Estimates	47
14.3.2	Quantitative Estimates	48
14.3.3	Tabulation of Categorical Estimates	48
14.3.4	Tabulation of Quantitative Estimates	49
14.4	Guidelines for Statistical Analysis	49
14.5	Coefficient of Variation Release Guidelines	50
14.6	Quality Level Guidelines	51
15.0	Variance calculation.....	52
15.1	Importance of the variance.....	52
15.2	Variance and CV calculation	52
15.2.1	CV look-up tables	52
15.2.2	Computation of Coefficients of Variation.....	53
15.2.3	Sample Requirements	53
15.2.4	Design Effect, Sample Size, Population	53
15.2.5	Approximate Sampling Variability Tables.....	55
15.3	How to Use the CV Tables For Categorical Estimates.....	55
15.4	Examples of use of the CV tables and of the coding for the nominal estimates	57
15.5	How to use the CV tables to obtain confidence limits	58

15.6	How to Use the CV Tables to Do a T-test	60
15.7	Coefficients of Variation for Quantitative Estimates	61
15.8	Release Cut-offs for the NLSCY	61
15.9	SAS and SPSS macros to calculate the variance, using the Bootstrap weights.....	63
16.0	Direct Assessment	64
17.0	Analytic Issues.....	65
17.1	How a complex sample design affects analysis	65
17.2	Unit of analysis.....	65
17.3	Type of analysis: cross-sectional, longitudinal or repeated	65
17.3.1	Longitudinal analysis	65
17.3.2	Cross-sectional analysis	65
17.3.3	Repeated analysis.....	66
17.4	Partial and total non-response.....	66
17.4.1	Total non-response.....	66
17.4.2	Partial non-response.....	67
17.5	Data processing	67
17.6	Computing the variance with certain software applications	67
18.0	Appendix I - Record Layouts.....	69
18.1	Education File	69
18.2	Secondary File	75

1.0 Introduction

The National Longitudinal Survey of Children and Youth – Education Component (NLSCY) for Cycle 4 was collected from April to June 2001 by Statistics Canada in partnership with Human Resources Development Canada. The Education Component is one portion of the second release of Cycle 4 data. The other portion of the release includes the individual variables used to create the scores for the Ages & Stages administered to children aged 3 to 71 months.

This manual has been produced to facilitate the manipulation of the Education microdata file, to document data quality and to explain other analytical issues. The first release of cycle 4 data was in June of 2003 and is often referred to throughout this manual as a resource for additional information.

This will be the last time that data will be collected from teachers and school principals for the NLSCY.

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

Statistics Canada

Client Services
Special Surveys Division
Telephone: (613) 951-3321 or toll free: 1-800-461-9050
Fax: (613) 951-4527
Email: ssd@statcan.ca

Special Surveys Division, Statistics Canada
2500 - R, Main Building
Tunney's Pasture
Ottawa, Ontario K1A 0T6

Human Resources Development Canada

Child, Youth and Social Development Studies
Applied Research Branch
Human Resources Development Canada
Place du Portage -Phase II, 7th floor
165 Hôtel de Ville Street
Hull, Quebec
K1A 0J2

Telephone: (819) 953-3465
Facsimile: (819) 953-8868

2.0 Background

The National Longitudinal Survey of Children and Youth (NLSCY) is a long-term study of Canadian children that follows their development and well-being from birth to early adulthood. The NLSCY began in 1994 and is jointly conducted by Statistics Canada and Human Resources Development Canada.

The study is designed to collect information about factors influencing a child's social, emotional and behavioural development and to monitor the impact of these factors on the child's development over time.

The survey covers a comprehensive range of topics including the health of children, information on their physical development, learning and behaviour as well as data on their social environment (family, friends, schools and communities). Information from the NLSCY is being used by a variety of people at all levels of government, in universities, and policy-making organizations.

The Education Component of the survey contains three separate questionnaires not included in the Main NLSCY component: a school principal questionnaire, a single teacher questionnaire and a multiple teacher questionnaire.

During the collection of Cycle 4 of the NLSCY, Statistics Canada experienced a heavier than anticipated interviewing workload due to the number of surveys in the field at the same time. Two months into the process, it became obvious that the collection of the school component was in jeopardy. A number of contingencies were considered and the following decisions were made: the math assessment would be collected at the household; the reading assessment would be dropped to reduce respondent burden and minimize costs; and, the school component would be limited to only the sub-sample of respondents where the school information could be collected and processed in time for the mail-out to schools before the end of the school-year. This last decision resulted in a lower return rate for the teacher and school principal questionnaires. To compensate for the non-response, a weight was created specifically for the Education file. This weight is for cross-sectional estimation and is the only weight that will be provided on the Education file. For more information about weighting please refer to Chapter 12.

Survey Population

In Cycle 4, a representative sample of Canadian children aged 0 to 17 years was followed for longitudinal and cross-sectional purposes. The Education Component includes only the children aged 6 to 15 who are attending public school. Note: there are 5 cases where the parent reported that the child was attending a public school but the school principal reported that the school was a private school.

Target population

The NLSCY objectives are to produce longitudinal and cross-sectional estimates. Therefore, several populations are targeted in the Cycle 4 sample.

- Cross-sectionally, the Cycle 4 sample represents all children who were aged 0 to 17 years on January 1st, 2001.
- Longitudinally, we have 3 cohorts:
 - o The first cohort represents all children who were aged 0 to 11 years in 1994/95. Those children are now aged 6 to 17 years in Cycle 4.
 - o The second cohort represents all children who were aged 0 to 1 years in 1996/97. Those children are now aged 4 to 5 years in Cycle 4.
 - o The third cohort represents all children who were aged 0 to 1 years in 1998/99. Those children are now aged 2 to 3 years in Cycle 4.

- The fourth cohort represents all children who were aged 0 to 1 in 2000/01. Those children will be followed until the age of 5.

The target population for the education file differs slightly from the main component. While the main component of the survey aims at representing all children in the respective age cohorts, the education files focus only on the children attending public school.

Collection Cycles

Data collection occurs at two-year intervals.

Cycle	Collection Start	Collection End
1	February 1995	April 1995
2	April 1997	June 1997
3	April 1999	June 1999
4	April 2001	June 2001

3.0 Objectives

The objectives of the NLSCY are:

- To determine the prevalence of various risk and protective factors for children and youth.
- To understand how these factors, as well as life events, influence children's development.
- To make this information available for developing policies and programs that will help children and youth.
- To collect information on a wide variety of topics – biological, social, economic.
- To collect information about the environment in which the child is growing up – family, peers, school, community.

Information comes from different sources (parent, child and teacher) and from direct measures (PPVT, math/reading tests, etc.)

Data Release Strategy

The Main microdata file for Cycle 4 was released in June of 2003 in conjunction with a re-release of updated files for Cycles 1, 2 and 3. The Education microdata file is a part of a secondary Cycle 4 release that will also include a separate microdata file containing the variables used to calculate the Ages and Stages scores (see Main file User Guide for details on this measure).

Ages and Stages Data Quality

It is important for data users to note that the non-response rates were particularly high for many of the Ages and Stages questions. When facing non-response, users should perform a non-response analysis before undertaking their analysis to assess the impact of the non-response. This is particularly important when the non-response is high. However, if the non-response is too high, it might become very difficult, sometimes impossible, to produce reliable estimates. As a rule of thumb, you should have a response rate of 67% or higher. With a response rate lower than 67%, adjusting for the non-response can become more problematic.

The Data Liberation Initiative (DLI) provides Canadian academic institutions with affordable access to Statistics Canada data files and databases for teaching and research. The cycle 4 files will be included on the DLI website at the following location:

[<<http://www.statcan.ca/english/Dli/Data/Ftp/nlscy.htm>>](http://www.statcan.ca/english/Dli/Data/Ftp/nlscy.htm).

4.0 Concepts and Definitions

There are many variables and concepts that are critical to the analysis of NLSCY data. The following is an explanation of the key concepts in the NLSCY. Derived variables are those that are not asked directly to the respondents but are calculated using information they have provided.

The content for each section of the various questionnaires used in the NLSCY is presented in the Main User Guide.

The unit of analysis for the NLSCY is the child or youth. See Chapter 5 for information on the survey design.

Only the concepts relevant to the Education file will be included here. Information related to concepts found in the Primary file can be found in the Main User Guide.

4.1 Definitions

Components

The NLSCY is made up of various components; these are generated by the computer application based on the child's age. The main components are: Adult, Child, Youth, Self completes, PPVT, Direct Measures, Math tests and Cognitive Measures. The Education questionnaires are an additional component asked of teachers and school principal of children whose parent and school board gave consent and who attend a public school (non-sectarian or sectarian/separate school).

Effective Age (DMMCQ01)

The age of the child determines, in most cases, the questions that will be asked about him or her. Instead of using the child's actual age, the NLSCY uses a calculated age called effective age. This is done to ensure the child stays in the age group to which he/she is assigned regardless of whether collection takes place before or after the child's birthday. For Cycle 4, the effective age is calculated as 2000 minus year of birth. For example a child born in 1998 would have an effective age of 2 years (2000 - 1998). Note that the actual age of the child at the time of the interview is sometimes different from the effective age.

Person Most Knowledgeable and spouse

In each NLSCY household, for each selected child, a question was asked about who in the household was the person most knowledgeable about this child. This person was labelled as the *Person Most Knowledgeable (PMK)*. The PMK provides the information for all selected children in the household and then gives information about himself/herself and his/her spouse/partner. In some rare cases it might have been appropriate to label two different people in a household as PMKs.

For example, in the case of a step family, it may have been appropriate to label the mother as the PMK for one child and the father for another. However, in order to simplify the interview procedures, only one PMK was selected per household.

In some households, there is no PMK. In cases where the selected child is 16 and over and is no longer living with a parent or guardian, there is no PMK selected in the household. The following is the breakdown of the relationship of the PMK to the NLSCY children for Cycle 4.

For 89.8% of responding children, the PMK was the mother (88.6% the biological mother and 1.2% the step, adoptive or foster mother)

For 8.7% of the children the PMK was the father. For the remaining 1.5% of children the PMK was not a parent.

For the majority of cases of the PMK not being a parent, the child had a parent living in the household, but the parent was not selected as the PMK. For the most part this situation occurred when a child had a very young mother living with her own parents, i.e., the child's grandparents, and the grandmother was selected as the PMK.

If the PMK had a partner residing in the household at the time of the interview, then this person was labelled as the spouse. Spouses included both married and common-law partners. Detailed socio-economic information was collected about the spouse/partner in order to describe the family situation of the child

Change in PMK between cycles

For several reasons, the PMK and his/her spouse could be different people than those designated in the previous cycle. For this reason, analysts should use caution when comparing PMK information from one cycle to the next.

4.2 Family Derived variables

For additional information about these variables please refer to the Main User Guide.

- Single Parent Families
- Intact, Step and Blended Families
- Siblings

4.3 Respondent

A cross-sectional respondent is a child from whom the Adult, Child or Youth component was completed. These children represent the population as of December 2000.

A respondent household is a household where an Adult component or a Child or Youth component has been completed.

Once the response status was determined at the household level, respondents were selected from all sub-sampled and eligible children on the education file. The following conditions had to be met for this selection: the child must be attending a public school (non-sectarian or sectarian/separate school) from whom we received parental consent, school board consent, and either a teacher (single or multiple) or a school principal questionnaire. These respondents were then assessed for their resemblance to the targeted population.

5.0 Survey Methodology - Sample

The NLSCY sample is a probability-based sample design that piggybacks to a large degree on the sample design of the Labour Force Survey (LFS). All respondents on the education file originate from the initial longitudinal panel of the NLSCY, were introduced into the sample in 1994 and were kept until the current collection cycle. The details of the full sampling methodology can be found in the main users guide.

As explained in the previous chapter, the respondents on the education file are a sub-sample of the eligible NLSCY sample. Inclusion in the sub-sample was based on the respondents' completed status at the time of the deadline for eligibility and, therefore, does not represent a true probability-based sub-sampling. This type of sampling is often referred to as "cut-off" sampling. Even though for each element we can calculate the probability of selection, there is a deliberate exclusion of part of the target population.

The sub-sample of eligible children represents about 70% of the total sample. This smaller sample and the reduced burden on the collection process for the school component did have a positive effect on the number of respondent returns. A full analysis of coverage issues arising from sub-sampling and an analysis of non-response was performed to assess the potential for bias. For full details, please refer to Chapter 10.

Readers should note that, although every effort was made for iterative proportional fitting of the sub-sample to the characteristics of the full sample, certain biases may not be fully accounted for by the new weights. For more information about weighting please refer to Chapter 12.

6.0 Data Collection

Data for Cycle 4 of the NLSCY were collected between the fall of 2000 and the spring of 2001. They were collected in two main settings: households and schools.

6.1 Household collection

The survey combines computer-assisted interviewing methods and the use of paper questionnaires. Depending on the composition of the household and the nature of the required components, the interview was conducted partly or completely by telephone and/or field visit.

This section provides a brief description of the “Collection tools” or the “Survey instruments”, in other words the computer-assisted and paper questionnaire components used in NLSCY collection.

6.1.1 Child component

A child component was created for each selected child between 0 and 17 years of age. The Person Most Knowledgeable (PMK) about the children and youth answered the child component questions. The PMK was usually the child’s mother, but it could also be the father, a step-parent or an adoptive parent who lived in the same dwelling. Only the PMK or his/her spouse was permitted to answer the questions in this component.

At the end of this component, the respondent was asked to provide the name, address and telephone number of two people (friends, relatives) who would be able to help us trace the family in two years, when the survey will be repeated.

For the children in primary school (aged up to 15 years, excluding those in kindergarten), in preparation for data collection in the schools, the PMK was asked to consent to the collection of information from the child’s teacher and school principal.

List of subjects covered

- Education
- Health
- Medical and biological information
- Mother’s work after the child’s birth
- Child’s development
- Temperament
- Literacy
- Communication
- Activities
- Behaviour
- Positive behaviour
- Sleep habits
- Motor and social development
- Relationships
- Parenting
- Custody
- Expectations (Aspirations)
- Socio-demographic characteristics

The following components are also part of the NLSCY and are described in detail in the Main User Guide:

- Adult component
- Youth component
- Peabody component (revised Peabody Picture Vocabulary Test – PPVT-R)
- Number Knowledge Component
- Control screen component
- Ages and Stages questionnaires
- Who am I? questionnaire for 4- and 5-year-olds
- Mathematics test (computation exercise)
- Self-complete Questionnaires – ages 10 to 11, 12 to 13, 14 to 15 and 16 to 17 years
- Cognitive test for 16- and 17-year-olds

6.2 School collection

This phase of data collection took place in the schools between April and June 2001.

For children in the sample who were attending school at a level higher than kindergarten and aged 15 or less, the PMK was asked to consent to the collection of information from the child's teacher and principal. School collection involved three questionnaires, which were mailed out to teachers and principals. They were asked to complete the questionnaires and mail them back to Statistics Canada in the envelopes provided. The package was to be sent to the principal, who was asked to give the questionnaire to the child's teacher. If the child had more than one teacher, the principal was asked to give the questionnaire to the teacher who knew the child the best, preferably a language or math teacher.

6.2.1 Teacher's questionnaire

Two questionnaires were developed, a first one for cases where the child had one teacher for the basic academic subjects and a second one for cases where the child had different teachers for the basic academic subjects. To determine which questionnaire to send to the teacher, parents were asked if their child had the same teacher for most of the basic subjects in school (i.e. mathematics, language arts and science). These questionnaires dealt with the child's academic performance and behaviour at school, the teacher's methods of instruction and the atmosphere in the classroom.

6.2.2 Principal's questionnaire

This questionnaire collected information about the teaching methods used in the school, the availability of educational resources, and the social atmosphere in the school. Hence, the Principal's Questionnaire was about school policies and the educational environment and not about a specific child.

7.0 Data Processing

7.1 Editing

The main output of the NLSCY is a "clean" master data file. This section presents a brief summary of some of the processing steps involved in producing this file.

The editing of the paper questionnaires was done using the Generalized Processing Environment. This is a generic system that the following series of steps to "clean" a file from beginning to end:

CleanUp
Pre-Edit
Flow Edit
Coding
Consistency Edit
Derived Variables
Final Processing File

7.2 Data Capture

Data Capture for Paper Questionnaires

Data capture for the Principal and Teacher questionnaires was accomplished at a centralized area at Statistics Canada's Head Office.

Any document containing at least one respondent-completed item was captured and a file containing each record was provided to Head Office processing staff for further processing. As part of the capture system, some quality checks were built in to flag unusual entries to warn the operators of potentially incorrect entries.

In cases where more than one response was checked off by the respondent, the operators were instructed to accept the first response. Errors remaining within the questionnaires were then edited at a later stage.

7.3 CleanUp

Defining Requirements

The purpose of this step is to drop full duplicate records and split off duplicate ID records for examination. Finally, split the data between response and non-response, with the given criteria.

A review was done of the responding and non-responding questionnaires and specifications were created based on this analysis to determine which records would be dropped due to non-response. Essentially, if a record was missing key information or had more than half the questions unanswered they were dropped from the file.

7.4 PreEdit

For all records where values were missing (blank), the value of "9, 99, 999..." was inserted to indicate that no information was collected. As well, the "Mark all that apply" responses were de-stringed to become yes/no questions. Finally, all text answers were removed from the processing file and set aside to be handled separately.

7.5 Flow Edits

The Flow Edits replicates the flow patterns from the questionnaire. Variables which are skipped based on flows are converted from *Not Stated* to *Valid Skip*.

Decision Tables are used to specify the flow edits. LogiPlus is used as the software to input the Decision Tables and generate the SAS code. A report with the Before and After counts of the variables is generated. Additionally, a report is generated providing the rule counts for each Decision Table.

For skips based on answered questions, all skipped questions are set to *Valid Skip*. For skips based on *Don't Know* and *Refusal*, all skipped questions are set to *Not Stated*.

7.6 Coding

The Coding process assigns standard codes such as NAICS to write-in fields, so that the data can be better analyzed. Also, *other specify* fields are examined and re-coded.

The coding was done manually by processing staff within the division. Questions which contain a list of answer categories often contain "*other specify*" as the final category. The text from the "*other specify*" question is captured. These write-ins are examined and may be re-coded up into one of the existing categories. If the write-in is reflected in one of the existing categories to the question, the appropriate category is set to *Yes* and the "*other specify*" is set to *No*.

7.7 Consistency Editing

The Consistency Edits process examines consistency between different variables. The first step was to match the Principal and Teacher questionnaires with respondents in the Main survey to ensure that they belonged in the sample. Also, any records where the parent or school board had not consented to participate in the survey were removed.

The second step involved comparing the reported grade of the child. In cases where there was a discrepancy from the parent-reported grade and the teacher-reported grade, the teacher-reported grade was accepted as correct.

The third and final step was to ensure consistency between questions within the actual questionnaire. Where responses were inconsistent the first answer was accepted as "correct".

7.8 Naming Convention and Coding Structure for NLSCY Variables

The NLSCY microdata file documentation system has employed certain standards to label variable names and values. The intent is to make data interpretation more straight-forward for the user.

Naming Convention for Variables

A naming convention has been used for each variable on the NLSCY data file in order to give users specific information about the variable. All variable names are at most eight characters long so that these names can easily be used with analytical software packages such as SAS or SPSS.

Format for Variable Names from Education File

“D” refers to the NLSCY Cycle 4

“EDU” refers to Education file

“T” refers to questions from both teacher questionnaires

“S” refers to questions from the single teacher questionnaires

“M” refers to questions from the multiple teacher questionnaires

“P” refers to questions from the school principal questionnaires

Format for Variable Names from Primary and Secondary Files

D SE C Q nnx or D SE C d Q nnx

“D” refers to the NLSCY Cycle

“A” indicates the first cycle,

“B” the second cycle,

“C” the third etc...

SE - refers to the section of the questionnaire where the question was asked or the section from which the variable was derived.

C - refers to the collection unit or the unit to which the variable refers.

There are five possibilities¹ :

“C” if the variable refers to the child,

“P” the PMK.

“S” the spouse/partner

“H” the household

“Y” if the variable refers to the youth,

d - the lower case letter refers to the NLSCY Cycle in which the variable first appeared on the file or cycle in which changes to a previously asked question were made.

Example: “b” indicates the variable was new in Cycle 2. In subsequent cycles, new variables will also be identified using the lowercase letter representing the cycle. New variables in Cycle 3 will contain a “c”, in Cycle 4 a “d”, etc. Some revisions were made to the content of the questionnaire between cycles. If the revision resulted in a change to the meaning or the values of a question, the variable was treated as new and contains a “d”.

Q refers to the variable for a question that was asked directly on one of the NLSCY questionnaires

“S” refers to a score calculated for one of the scales used on the questionnaire

“D” means the variable was derived from other questions that were asked on the questionnaire

“I” means the variable is a flag created to indicate that an item has been imputed

¹It should be noted that while variables do exist for various units of analyses (i.e., the PMK, the spouse/partner and the household), it will only be possible to produce “child estimates” from the NLSCY file. The characteristics of the PMK, spouse/partner and household can be used to describe attributes of the child. For example it will be possible to estimate the number of children living in a household with low income, or the number of children for whom the PMK has scored high on the depression scale etc. However it will **not** be possible to produce estimates of the number of low income households or depressed PMKs.

“Z” means the variable is a flag created to indicate an inconsistency in reported data between the current and previous cycles

“nnx” refers to the question or variable identification. Generally nn is a sequential number assigned to the variable; and x is a sequential alphabetic indicator for a series of variables of a similar type

7.9 Acronym Names for Questionnaire Sections

The following table gives the acronyms that were used for each section of the various NLSCY questionnaires. This acronym is embedded in the variable name for all variables on the NLSCY data file. The acronym is the second and third characters of the variable name.

	Variable	Collected or Derived from the:
GE	Geographic	sample information
HH	Household	dwelling characteristics
MM	Variables collected as part of the household roster.	Basic demographic variables for each household member. These variables are included on the NLSCY data file for the child, the PMK and the spouse/partner
DM	Demographic- derived to explain the living arrangements of the child:	information of the household roster and relationship grid
SD	Socio-demographic	child on the Child's Questionnaire and for the PMK and spouse/partner on the Adult Questionnaire.
HL	Health	PMK and Spouse on the Adult questionnaire, and for the Child on the Child questionnaire, and for the Youth on the Youth questionnaire
CH	Adult Chronic Conditions	PMK and Spouse in the Health section of the Adult questionnaire
RS	Restriction of Activities :	PMK and Spouse in the Health section of the Adult questionnaire
DP	Depression scale	Parent Questionnaire (this scale was administered to the PMK)
ED	Education	children aged 4 to 15 years on the Child's Questionnaire and about the PMK and spouse/partner on the Adult Questionnaire and for the Youth on the Youth questionnaire
LF	Labour force	PMK and spouse/partner on the Adult Questionnaire and for the Youth on the Youth questionnaire
IN	Income	household income and personal income of the PMK, collected on the Adult Questionnaire and for the Youth on the Youth questionnaire
FN	Family functioning	Adult Questionnaire (section asked to the PMK or spouse)
MD	Medical/biological	Child's Questionnaire (aged 0 to 3 years)
TM	Temperament	Child's Questionnaire (3 months to 35 months)

LT	Literacy	Child's Questionnaire (0 to 9 years)
AA	Activities	Child's Questionnaire (0 to 15 years) and for the Youth on the Youth questionnaire
BE	Behaviour	Child's Questionnaire (0 to 11 years)
MS	Motor and social development	Child's Questionnaire (0 to 47 months)
RL	Social relationship	Child's Questionnaire (4 to 9 years)
PR	Parenting style	Child's Questionnaire (0 to 15 years)
CR	Child care	Child's Questionnaire (0 to 13 years)
PP	PPVT test:	aged 4 to 6 years (if child in grade 1 or less included those over 6 years of age)
FF	Friends and Family	10 to 17 Self-complete Questionnaires
SC	School	10 to 17 Self-complete Questionnaires
AM	About Me	10 to 17 Self-complete Questionnaires
FB	Feelings and Behaviour	10 to 17 Self-complete Questionnaires
PM	My Parents and Me	10 to 17 Self-complete Questionnaires
PU	Puberty	10 to 17 Self-complete Questionnaires
DR	Smoking, drinking and drugs	10 to 17 Self-complete Questionnaires
AT	Activities	10 to 17 Self-complete Questionnaires
HT	Health	10 to 17 Self-complete Questionnaires
WK	Work and Sources of Money	10 to 17 Self-complete Questionnaires
DA	Dating	10 to 17 Self-complete Questionnaires
MA	Math computation test	children in grade 2 and over and Cognitive Math test for 16 to 17 year olds.
SF	Neighbourhood Safety	Adult Questionnaire (section asked to the PMK or spouse)
SP	Social Support	Adult Questionnaire (section asked to the PMK or spouse)
SL	Sleep	Child's Questionnaire (0 to 13 years)
PB	Positive Behaviour	Child's Questionnaire (3 to 5 years)
AS	Aspirations	Child's Questionnaire (16 to 17 years)
AG	Ages & Stages	Parent Questionnaire (3 to 71 months)
WM	Who am I	Direct Measure (4 to 5 years)

KN	Number Knowledge	Direct Measure (4 to 5 years)
OB	Observation Assessment	All children who receive Direct Measure
TS	Single Teacher	Single Teacher Questionnaire
TM	Multi-Teacher	Multi-Teacher Questionnaire
PC	Principal	School Principal Questionnaire

7.10 Examples of Variables Names

In order to illustrate the naming convention used for variables included on the NLSCY data file the following examples are given.

Variable Name Refers to:

Variable Name	Refers to :
DLFSQ2	Q2 in the Labour Force Section for the spouse/partner
D	Cycle 4 variable
LF	the Labour Force Section
S	The spouse/partner
Q	An item asked directly on the questionnaire
2	the ID of the item

Variable Name	Refers to :
DPRCS03	a positive interaction score on the parenting scale for a 2 to 15 year-old child
D	Cycle 4 variable
PR	the Parenting Section
C	the child
S	a score
03	the ID of the item

7.11 Coding Structure for NLSCY Variables

Some standards have been developed for the coding structure of NLSCY variables in order to explain certain situations in a consistent fashion across all variables. The following describes these various situations and the code used to describe the situation.

Refusal

On the Main data file an item which was refused is indicated by a code "8". For a variable that is one digit long the code will be "8", for a 2 digit variable "98", for a three digit variable "998" etc. Since the responses are self-reported on a paper questionnaire, this particular code is not on the variables from these questionnaires.

Don't Know

The respondent may not know the answer to a particular item. On the Education data file, the code used to indicate that the respondent did not know the answer to an item is "7". For a variable that is one digit long the code will be "7", for a two-digit variable "97" for a three-digit variable "997" etc.

Not Applicable

In some cases a question was not applicable to the survey respondent. A code "6", "96" "996" ... has been used on the data file to indicate that a question or derived variable is not applicable.

In some cases a single question or series of questions was not applicable. For example, the question on styles for teaching reading, only teachers of children who are being taught this subject are included. Otherwise there will be a code "6" for this question.

Not-Noted

In some cases, as part of Head Office processing the answer to an item has been set to not-noted. The not-noted code indicates that the answer to the question is unknown. Not-noted codes were assigned for several reasons.

- In some cases a questionnaire was not started or it was started but ended prematurely. For example, there may have been some kind of an interruption and the questionnaire was then mailed back partially complete. If there was enough information collected about the respondent, then all remaining incomplete items on the questionnaire were set to not-noted. The one exception was that if it was known that a certain section or a certain questionnaire was not applicable, then these questions were set to not applicable.
- An additional situation in which not-noted codes were used was as a result of consistency edits. When the relationship between groups of variables was checked for consistency, if there was an error, often one or more of the variables was set to not-noted.

For derived variables if one or more of the input variables to the derived variable had a refusal, don't know or not-noted code, then the derived variable was set to not-noted.

7.12 Imputation flags

Missing Values

Certain variables from the Education component could be missing and this is usually referred to as item non-response or partial response. In fact, certain respondents may not have responded to all the questions. This same situation also occurs with the main component.

Imputation

The Education file is made up of questions asked in both the teacher and the school principal components. If one of these components was answered and the other totally not answered, the component not answered was imputed in its entirety. However, in the event that both

components were entirely not answered, the record was considered to be non-respondent and no imputation was made.

Imputation involves determining the probable answers for missing responses (see chapter 11). The two data flags for the micro data file for the NLSCY Education component were FLAG_TE and FLAG_PR. FLAG_TE = 1 indicates that the teacher data (TEacher) were imputed. Its value was 0 otherwise. The FLAG_PR variable = 1 indicates that the Principal data were imputed. The variable was valued at 0 otherwise.

Please note that if one or several of the variables of an answered component presented a non-response, this or these variables were not imputed.

7.13 Derived Variables

Combining Items

A number of data items on the data file have been derived by combining items on the questionnaire in order to facilitate data analysis. For example, the questions about years and months of experience as a teacher/principal was combined into one variable indicating the total number of months of experience reported.

Derived Variable Name

All derived variables on the NLSCY data file have a "D" as the fifth character of the variable name.

8.0 Content of the Survey

The NLSCY was designed to follow an ecological or holistic approach to measuring child development. The survey captures the diversity and dynamics of the factors affecting children. To ensure that all relevant topic areas affecting child development were adequately addressed by the survey, a multidisciplinary consultation was carried out at the inception of the survey. The selection of specific subject areas, priorities and survey questions was very much a group effort with input and advice from:

- the NLSCY expert advisory group that consists of researchers in the area of child development and the social sciences;
- federal departments;
- representatives from the provinces and territories responsible for child development programs.

It was recommended that the NLSCY cover a broad range of characteristics and factors affecting child growth and development. Extensive information was gathered about the child, as well as the child's parent(s), characteristics of the family and the neighbourhood as well as the child's school and school experiences. This section provides an outline of the content for each section of the questionnaire included in the NLSCY data.

8.1 Processing System

As part of the NLSCY processing system, there were some basic quality checks performed for each section of the questionnaire. Any items for which there was a high level of non-response were looked at in detail. Any concerns about potential data quality problems for any items in a particular section of the questionnaire are discussed in this section of the document.

For a discussion on the validation of the scale scores, please see Chapter 9.

8.2 Demographic Variables

The demographic variables are collected in the household questionnaire. As part of the household questionnaire some basic demographic information (e.g., age, gender, and marital status) is collected for all members of the child's household. The relationship grid is also completed as part of this questionnaire i.e., the relationship of everyone in the household to the PMK. Using this information it is possible to create an extensive set of variables to describe the child's family situation. Most of these derived variables are critical to the analyses of NLSCY data and are described in the Main User Guide.

8.3 Adult Questionnaire

Below are the components of the Adult questionnaire, please refer to the Main User Guide for a more detailed description.

Education (Parent) - The Education Section is completed for both the PMK and spouse/partner. The objective is to gather information on the years of school completed, educational attainment, and current attendance at an educational institution.

Labour Force - The Labour Force Section is completed for both the PMK and spouse/partner. The main objective of the section was to determine employment stability as an indicator of the continuity of employment income.

Income - In the Income section of the survey, the sources of income and the income are collected for each household. Income range is also collected for the PMK and for the PMK's spouse. This information provides an indicator of the family's economic situation, an essential component of the child's environment.

Health - This section asks the PMKs and their spouses about general health, chronic conditions, restriction of activities as well as questions on smoking and drinking. The smoking questions have been included because research has indicated that parental smoking behaviours may be predictive of the use of cigarettes by children. Alcohol consumption is covered because of potential impacts on the adult's physical or mental health, the family's economic situation, and family relationships.

Neighbourhood Safety - This section gathers information about the respondent's satisfaction with his/her neighbourhood as a place to raise children, including perception of the extent of danger and problems, and of social cohesion or "neighbourliness".

Socio-demographic Characteristics - The objective of the Socio-demographic Section is to gather information on immigration, ethnic background and the language profile of household members. This will allow for analysis of various components of the Canadian population and will permit identification of visible minorities. As well, there is a question on religious affiliation. These questions are asked of PMK, spouse and the child.

Due to a technical problem, no socio-demographic questions were asked to returning respondents in Cycles 3 or 4. The data on the file reflects the respondent's answers in Cycle 2. Users should be aware that information about language(s) spoken, in particular, may have changed since we last asked these questions and should use these variables with caution.

8.4 Child Questionnaire

Education (Child) - Basic information is collected for all age groups, such as: the child's grade level, type of school and language of instruction, whether the child looks forward to school, absenteeism, number of school changes and residential moves.

For children in grade 1 or higher, additional questions are asked concerning other aspects such as skipping and repeating grades, achievement and special education.

Health (Child) - The objective of this section is to provide information on the child's physical health – general health, injuries, limitations and chronic conditions – and use of health services and medications.

Medical/Biological - The Medical /Biological Section was completed for children in the zero to three age group. The major objective is to collect information on factors such as gestational age and birthweight. These factors have been shown to have a direct impact on a child's growth and development. For example, in the long term, underweight babies face higher risks of poor health as well as longer-lasting developmental difficulties.

Work After Birth - These questions are asked to determine the time interval after which mothers returned to work following the birth of a child and the extent to which these mothers participated in the labour force upon their return.

Milestones - These questions were added in Cycle 4 to provide a better measure of early child development. Taken as a package, developmental milestones, such as when the child first said words or took first steps, provide a general sense of a child's development.

Temperament - This section measures the temperament of young children by asking the parent about the degree of difficulty their child presents them. This measure is based on the assumption that a child's temperament is influenced by the parent's perception of the difficulty of the child, and that temperament is not solely based on biological origins.

Literacy - This section measures children's exposure to books and their interest in reading and learning-related activities that parents do with their children. The focus of this section is the stimulation young children receive at home.

Communications - The items have been modified from the New Zealand Competent Children Project. They cover a child's ability to understand oral messages and to pass a message on to someone else, as well as to communicate verbally. The final question, about speech being easily understood, is only asked of three-year olds.

Activities - This section measures the child's participation in various non-school activities and the amount of household responsibility taken on by 10- and 11-year olds at home. The latter questions are used to create the Home Responsibilities Score (DACCS6), indicating the degree of home responsibilities. The section will give some sense of how the child spends his/her time, of personal interests, as well as the degree of interaction with peers.

Behaviour - The objective of this section is to assess aspects of the behaviour of children aged two years and older and of feeding patterns for 1- to 3-year-olds. The questions in this section are used to measure the prevalence of behaviours such as hyperactivity and physical aggression.

Positive Behaviour - The objective of this new section is to assess positive behaviour of children aged three to five, including perseverance and independence. The New Zealand's Competent Children Project has found that perseverance and independence were among a cluster of competencies that are good indicators of a child's overall performance. Questions have been adapted from the New Zealand study and the behaviour questions used for other ages in the NLSCY.

Sleep - Research suggests that sleeping difficulties are predictive of a child's potential difficulties. Conversely, absence of such difficulties has been correlated with easy temperament and positive outcomes. Several new questions, adapted from the Étude longitudinale du développement des enfants du Québec (ELDEQ) – Santé Québec, were added in Cycle 4. These questions asked about hours of sleep, hours of uninterrupted sleep at night, how often the parents sleep was disturbed by the child and so on.

Motor Social Development - The Motor and Social Development (MSD) Scale measures dimensions of the motor, social and cognitive development of children from 0 to 3 years; the questions vary by the age of the child. Three scores (DMSCS01, DMSCS02 and DMSCdS03) are derived from these questions.

Relationships - The objective of this section is to provide information about the child's relationships with others. Positive relationships with other children and adults may help to counteract other factors that place a child at risk. Questions about doing things with friends, and getting along with parents, teachers and friends are based on those in Ontario Child Health Survey.

Parenting - Parenting style is considered to have an important influence on child behaviour and development. The objective of this section is to measure certain parenting behaviours. Scales are created from the questions in this section.

8.5 Cautionary Note

Due to low return rates, the Cycle 4 Education file was re-weighted unlike in previous cycles where the weights from the primary file were included. A cross-sectional weight was created that compensated for the reduced sample. Analysis was done using a selection of variables from the Main component that have been included on the Education file. By doing proportional fitting for these selected variables, an adjusted weight was created.

By including variables from the main component of this file it is hoped that users will be able to do most of their cross-sectional analysis without the main component. Should data users require the main file, they should note that their estimates may differ for variables not used for calibration, since any coverage differential from the main and sub-sampled files would not be accounted for by the adjusted education weight. Similarly, any longitudinal analysis will require the main file in order to obtain the longitudinal weight; the education information will only apply to a subset of the entire file, with no adjustment to account for the missing information.

9.0 Validation of the survey scales

9.1 Validation of scale data

9.1.1 Scale Definition

For some of the concepts deemed important to measure in the NLSCY it was decided that the concept would most appropriately be measured through the use of a scale. A scale is simply a group of questions or items that measure a certain concept when the answers to the items are put together.

For example, on the child's questionnaire, it was determined that it was important to have an assessment of certain parenting behaviours. The scale is intended to measure three different constructs or factors related to parenting; positive interaction, ineffective parenting and consistent parenting.

9.1.2 Scales and Calculations

For each factor measured by a scale, a score is calculated. The score for a particular factor can be used to give an ordering of individuals. For example, for the Parenting Scales, for children with higher scores for the "positive interaction" factor, the PMK reported having more positive encounters with the child (e.g., laughed with them more, praised them more etc.). The score for a particular factor is usually based on a series of items, since one single item usually cannot measure the factor or construct with adequate precision.

During the development of the NLSCY, when consideration was being given to what scales should be used to measure a particular concept, an attempt was made to select scales that had been used in other studies. In this way, the psychometric properties of the measures produced by each scale were available with complete references.

9.1.3 Evaluation of Scale Data

In many instances, the wording of certain questions in the original scale was modified and in some cases new questions were added. Sometimes the scale that was used had not previously been used for children in Canada, or had only been used for very small samples. Given these concerns and further concerns regarding interviewing conditions, it was felt that the factor structures of the scales used in the NLSCY could be different from the ones given in the literature. Therefore the project team felt the need to carry out an extensive evaluation of the scale data to ensure that the psychometric properties found in other studies also held true for the NLSCY experience.

There were three major steps in the analyses of the scale data. First a new factor analysis was performed on all scales to determine the constructs or factors inherent in each scale. Then scale scores were calculated based on this factor structure. Finally reliability measures were produced. The general procedures followed for each of these steps are described in detail in the following pages.

NOTE: Many of the scales were developed and validated in Cycle 1. In subsequent cycles, the same factor structure which emerged from the Cycle 1 analysis was imposed. Imposing the same factor structure ensures that the scales are consistent

across time to allow for longitudinal analysis and cross-sectional comparisons. Each scale has a note indicating in which cycle the validation was performed.

9.2 Factor analyses

9.2.1 Factor Analysis for Scales

The factor structure of each scale was determined based on data from the first cycle. The factor structure imposed on the scales already used in the first cycle and repeatedly utilized in the second, third and fourth cycles of the survey was the result of analyses of data from the first cycle. This analysis was redone after cycle 4 and it is these results that are summarized below. For the results from the cycle 1 factor analysis please refer to the cycle 1 User Guide.

1. The sample of respondents for each scale (and age group, if the scale used different questions for different groups), was randomly divided into two half samples. This was done to find out whether different samples would yield the same results.
2. Principal component analysis was carried out separately on each half-sample to find out how many factors should be extracted in the subsequent factor analysis. In principle, the same number of factors as was found in the literature was expected. In practice, however, some scales showed a different number of factors because in some cases factors combined while in others new factors emerged.
3. Factor analysis was done on each half-sample and the factor structure and loading of each factor were compared across the half-samples.
4. In the factor analysis, the items for each child in the appropriate age group were used, multiplied by the child's normalized weight. An individual's statistical weight is normalized by dividing his/her weight (DWTCW01C) by the average weight for all individuals. Thus, the sum of the normalized weights is equal to the sample size.
5. Once the factor structures were analysed and the items included in each factor were determined, scores were calculated. To produce the scores, 1 was subtracted from each item so that the lowest possible score would be 0. A score of 0 indicates that the child has no problems for all factors in the behaviour scale except for the Prosocial factor, where a score of 0 indicates the absence of prosocial behaviour. Some items were imputed. The imputed values were computed by a procedure (the SAS PRINQUAL procedure) that determines which of the possible values for an item is the most plausible for an individual in view of his/her response profile, the response profiles of others in the sample, and the number of factors included in the analysis.
6. The score for each factor on the scale was derived by totaling the values of the items that made up that factor (including imputed values). The score was set to "missing" if too many of the values of any items included in the factor were unreported. A value may be missing if the parent refused to answer or did not know the answer to the item.

9.2.2 Data Transformation Using Optimal Scaling

Before performing the factor analysis for each of the NLSCY scales, the data were transformed using optimal scaling. The method used was one proposed by Young and several associates (Young, 1981) which is a variant of Fisher's optimal scaling technique. The method is presented as a means of transforming data that are fundamentally nominal or ordinal in nature to interval or ratio level data so that statistical techniques which are appropriately applied only to interval and ratio data may be utilized.

9.2.3 Factor Analysis Using Weighted Data

Factor analysis requires that the data have the property of interval or ratio data, meaning that the distance between each answer category of the question should be the same. For example, in scales where the answer choices are: Never, Sometimes, Often, and Always, one must assume that the distance between Never and Sometimes is the same as that between Sometimes and Often in the respondent's perception. It was felt that this was not necessarily true in the case for the scales used in the NLSCY and steps were taken to compensate for this difficulty (data transformation).

9.3 Calculation of scores and item imputation

9.3.1 Calculation of Scores for Each Factor

The results of the factor analyses were used to determine what items "loaded" into each factor (i.e. were a part of each factor). The next step was to calculate a score for each factor. This was done by summing the values for each individual item that made up the factor. In some cases some rescaling of values was done before the final score was calculated. The following example illustrates how factor scores were computed.

9.3.2 Example of Factor Score Computation

One of the constructs that emerged in the factor analysis for the Parenting Scale on the Child's Questionnaire was the ineffective parenting factor (Age d 2 to 11 Years). In the factor analysis on cycle 1 data, seven items were found to load into this factor.

DPRCQ04	How often do you get annoyed with your child for saying or doing something he/she is not supposed to?
DPRCQ08	Of all the times you talk to your child about his/her behaviour, what proportion is praise?
DPRCQ09	Of all the times you talk to your child about his/her behaviour, what proportion is disapproval?
DPRCQ13	How often do you get angry when you punish your child?
DPRCQ14	How often do you think the kind of punishment you give your child depends on your mood?
DPRCQ15	How often do you feel you have problems managing your child in general?
DPRCQ18	How often do you have to discipline your child repeatedly for the same thing?

The answer categories for these items were of two types:

- 1 - never
- 2 - about once a week or less
- 3 - a few times a week
- 4 - one or two times a day
- 5 - many times each day

- 1 – never
- 2 - less than half the time
- 3 - about half the time
- 4 - more than half the time
- 5 - all the time

In the calculation of the score for this ineffective parenting factor, the categories were rescaled to 0 to 4 (i.e., the category "never" was scored as 0, the category "about once a week or less/less than half the time" was scored as 1, ... and the category "many times each day/all the time" was scored as 4). In order to compute the score, these values were summed across the seven items involved in the factor resulting in an ineffective parenting score in the range 0 to 28. A low score of 0 represents a more effective parenting style and a high score of 28 indicates a more ineffective parenting style. For most of the scores calculated for the NLSCY, a score of 0 represents the absence of a problem. However there are exceptions to this which are noted in the documentation for each particular scale.

9.3.3 Negative Loading

Note that the second item that loaded into the ineffective parenting factor, DPRCQ08 (Of all the times you talk to your child about his/her behaviour, what proportion is praise?) is in the opposite direction compared to the other items. In fact the item loaded "negatively" into the factor. Therefore when computing the score the values for this item were reversed - all the time was scored as 0, more than half the time as 1, ... and never as 4. In the documentation for each scale any item that was reversed for the scoring algorithm due to a negative loading is indicated.

9.3.4 Non-Response Code

The score for the ineffective parenting factor is labelled as DPRCS04 on the record layout for the micro data file. An "S" in the 5th position of the variable name indicates a score.

When the score was being calculated for each factor there was a possibility that one or more of the items making up the score had a non-response code (don't know, refusal or not-stated). If the number of items with a non-response code was above a certain threshold, the factor score was set to not-stated. Generally this threshold value was set at 10% of the items.

9.3.5 Raw Items

It should be noted that in addition to the scores, the raw items for each scale are included on the micro data file. This will allow researchers to consider alternate factor structures if desired. For the raw items the original values (in the 1 to 5 range for the parenting scale) have been retained before any rescaling or reversal of values took place.

9.4 Reliability measures for scales

Reliability refers to the accuracy, dependability, consistency or ability to replicate a particular scale. In more technical terms, reliability refers to the degree to which the scale scores are free of measurement error. There are many ways to measure reliability.

9.4.1 Cronbach's Alpha

One of the most commonly used reliability coefficients is Cronbach's alpha (Cronbach, 1951). Alpha is a measure of the internal consistency of the items within the factor. It is based on the average covariance of items within the factor. It is assumed that items within a factor are positively correlated with each other because they are attempting to measure, to a certain extent, a common entity or construct.

9.4.2 Interpretations of Cronbach's Alpha

Cronbach's alpha has several interpretations. It can be viewed as the correlation between the scale or factor and all other possible scales containing the same number of items, which could be constructed from a hypothetical universe of items that measure the characteristic of interest. For example, in the ineffective parenting factor, the seven questions included in the scale can be viewed as a sample from the universe of all possible items. Parents could also have been asked: "How often do you raise your voice when you discipline your child?" or "How often do you threaten punishment more often than you use it?" Cronbach's alpha indicates how much correlation can be expected between the scale which was used and all other possible seven-item scales measuring the same thing.

Another interpretation of Cronbach's alpha is the squared correlation between the score an individual obtains on a particular factor (the observed score) and the score he/she would have obtained if questioned on all possible items in the universe (the true score). Since alpha is interpreted as a correlation coefficient, it ranges from 0 to 1.

Generally, it has been shown that alpha is a lower bound to the reliability of a scale of n items (Novick and Lewis, 1967). In other words, in most situations alpha provides a conservative estimate of a score's reliability.

9.4.3 What is a satisfactory level of reliability?

It is difficult to specify a single level that should apply in all situations. Some researchers believe that reliabilities should not be below 0.8 for widely used scales. At that level, correlations are affected very little by random measurement error. At the same time, it is often very costly in terms of time and money to obtain a higher reliability coefficient. It should be noted that for some of the factors for which scores were computed for the NLSCY, the reliabilities are below this level. The Cronbach's alpha is given in the documentation for each score that has been calculated. Researchers can determine for themselves whether or not the score has adequate reliability for their specific purposes.

Finally, it should be mentioned that for the NLSCY the Cronbach's alpha for each factor score was computed using SAS. Typically, the alpha coefficients calculated using SAS are lower than those calculated using SPSS.

9.5 Cycle 4 Education Scales

The scales included from the Main file are not described here; please refer to the Main user Guide for additional information.

The Teacher Scales

This particular section of the National Longitudinal Survey of Children Youth was designed to increase the knowledge of the impacts that educationally relevant variables have on child outcomes. Questions were administered to the teachers of every eligible school-aged child in the survey whose parents gave consent and where consent was received from the school boards. The various scales included in this section were developed by the Centre for Education Statistics (CES) of Statistics Canada in consultation with the NLSC Project Team and Drs. D. Willms and F. Echols of the University of British Columbia. It must also be mentioned that in the Social Skills Index Score, the list of social/personal skills and work habits, were taken from Freeman and Hatch's (1989) research into commonly mentioned comments on report cards.

The following section includes a list of the scales contained within the Teacher's Questionnaire of the NLSCY and a break down of which cycles used these scales and which individual questions have been added or removed. As the age of the children in the sample increased, it was necessary to create questions and scales that would reflect this. As such, in cycle 1, it was necessary to develop a kindergarten teacher's questionnaire. In cycle 2, the questionnaire was adapted to accommodate children who had one teacher for different academic subjects and one for children with different teachers for different subjects. To further increase the accuracy of the survey, in cycle 3, a separate questionnaire was given to the language and math teachers of children who had different teachers for different subjects. In cycle 4, a similar format to that used for cycle 2 was employed.

There were few changes made to the scales used in the Teacher's Questionnaire portion of the survey. The variable **DETKQ17G** (how often does the student demonstrate self-confidence) was added to the Social Skills Index for cycles 2 and 3. The "Engagement with School" score was re-titled "School Preparedness" for cycle 3. Finally, the Academic Press Scale was dropped altogether in cycle 3.

The Scores for the Various Teacher-Reported Scales

The following analysis was done using data from the third cycle of the NLSCY, collected in 2000 but the variable names begin with a "D" to reflect the current cycle.

Social Skills Index Score:

This score was derived using the following items: DETCQ17A, DETCQ17B, DETCQ17C, DETCQ17D, DETCQ17E, DETCQ17F, DETCQ17H, DETCQ17I and DETCQ17J. No imputation was done for this score. The score varies between 0 and 36, a high score indicating a high indication of social and personal skills demonstrated by the child in the class. The Cronbach Alpha value is 0.93.

Engagement with School Score/School Preparedness Score:

This factor was derived using the following: DETCQ26A, DETCQ26B, DETCQ26C, DETCQ26D and DETCQ26E. No imputation was done for this score. The score varies between 0 and 20, a high score indicating lower levels of preparedness for school. The Cronbach Alpha value is 0.83.

Conduct Disorder and Physical Aggression Scale

This factor was derived using the following: DETCQ27G, DETCQ27X, DETCQ27AA, DETCQ27FF, DETCQ27JJ, and DETCQ27NN. No imputation was done for this score. The total score ranges from 0 to 12, a high score indicating the presence of conduct disorder and physical aggression. The Cronbach Alpha value for this factor is 0.887.

Indirect Aggression

This factor was derived using the following: DETCQ27J, DETCQ27R, DETCQ27Z, DETCQ27LL, and DETCQ27SS. No imputation was done. The score varies between 0 and 10, a high score indicating the presence of indirect aggression. The Cronbach Alpha value for this factor is 0.886.

Participative School Environment

This factor was derived using the following: DETCQ50A, DETCQ50B, DETCQ50C, DETCQ50D, DETCQ50E, DETCQ50F and DETCQ50G. No imputation was done for this score. The score varies between 0 and 28, a high score indicating a high participative school environment for the teacher. The Cronbach Alpha value is 0.88.

Supportive School Environment

This factor was derived using the following: DETCQ50H, DETCQ50I, DETCQ50J, DETCQ50K, and DETCQ50L. No imputation was done for this score. The score varies between 0 and 20, a high score indicating a high supportive school environment. The Cronbach Alpha value is 0.91.

Disciplinary Climate Score

This factor was derived using the following: DETCQ51A, DETCQ51B, DETCQ51C and DETCQ51D. No imputation was done for this score. The score varies between 0 and 16, a high score indicating the presence of a strong disciplinary climate in the school. The Cronbach Alpha value is 0.81.

The Principal-Reported Scale

There was only one scale included in the Principal's portion of the questionnaire of the NLSCY. It was the **Disciplinary Climate Score**. This factor was derived using the following: DEPCQ15A, DEPCQ15B, DEPCQ15C, DEPCQ15D, DEPCQ15E, DEPCQ15F, DEPCQ15G, DEPCQ15H, DEPCQ15I, DEPCQ15J, DEPCQ15K and DEPCQ15L. No imputation was done for this score. The scores vary from 0 to 46, a high score indicating the presence of disciplinary problems in the school. No Cronbach's Alpha value was noted.

10.0 Survey methodology - Response rates and Coverage of the Education Component

Two types of non-response were evaluated in the Education component : total non-response where the selected respondent (teacher or school principal) did not answer any questions and partial non-response where the selected respondent (teacher or school principal) answered enough questions to be considered a respondent, but did not answer all the questions.

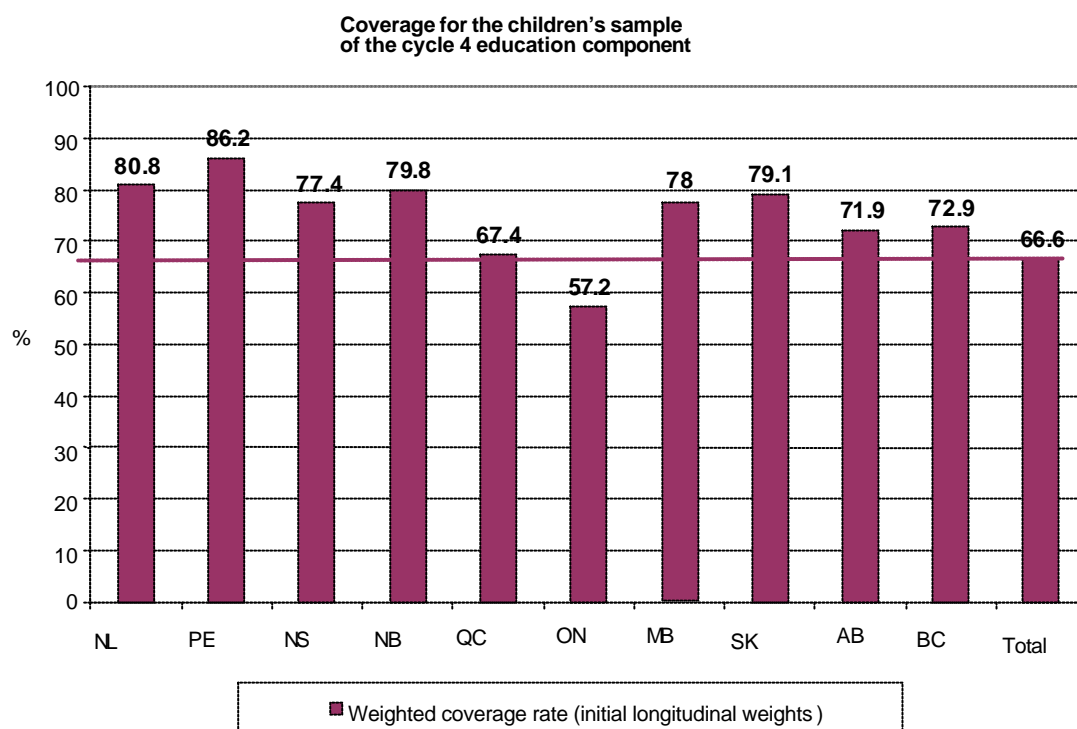
In addition to the non-response, we took into account the problem of under-coverage caused by operational difficulties during the collection.

This section deals with total non-response. More than 10,000 children were initially eligible for the survey. Operational problems reduced the sample and resulted in under-coverage for the NLSCY Education component for cycle 4. The non-response occurred at various points: the parents either did or did not consent to us contacting the school boards, the school boards themselves did or did not consent to us contacting the child's teachers and the principal of the child's school. Finally, the teachers and/or school principals may not have responded to the questionnaire relating to the child or the child's school.

Response rates in a survey are used to measure the effectiveness of the collection process and are also a good indicator of the quality of the estimates produced. The partial non-response rate is covered in the following section on imputation.

10.1 Survey undercoverage

Only a part (70%) of the eligible children (living in one of the 10 provinces, attending a public school, born between 1985 and 1994) were sampled because of the operational problems. This coverage was, however, different based on the province of residence or the size of the area of residence or finally the age of the child. There was no difference in coverage based on the child's sex.



Coverage rate by size of area of residence

Size of area of residence	Rate
rural region	68.8
urban region < 30,000	68.1
urban region 30,000 – 99,999	66.4
urban region 100,000 – 499,999	68.1
urban region more than 500,000	<u>63.7</u>
Not stated	74.7
<i>Total</i>	66.3

Coverage rate by age

Child's age	Rate
6 years	59.4
7 years	53.8
8 years	58.6
9 years	66.3
10 years	67.8
11 years	68.5
12 years	70.2
13 years	72.7
14 years	72.3
15 years	74.6
Total	66.3

10.2 Total Non-response

10.2.1 Definitions

Some definitions are needed in order to understand the tables in this chapter.

A **respondent household** is a household for which an adult component or a child or youth component was completed.

A **respondent child** is a child for whom an adult component or his child or youth component was completed. A respondent household that does not have a complete adult component may have one respondent child and one non-respondent child. There were 19 children in that situation.

An **out-of-scope household** is a household whose children are all deceased or living outside Canada's 10 provinces.

The **response rate** is defined as the number of respondents over the number of in-scope children.

Refusals are identified during the processing of response files. The interviewers' notes are used, and a keyword search is performed to complete the coding.

The **refusal rate** is defined as the number of refusals over the number of in-scope children.

A household is **to be traced** when a call made in connection with the case results in a trace response code.

The **to-be-traced rate** is defined as the number of cases to be traced over the total sample (in scope and out of scope).

A household is considered **not traced** when it is non-respondent and the final response code is a trace response code.

The **successful trace rate** is defined as the number of successfully traced children over the number of children to be traced.

Parental consent rate is the rate of non-response that refers to the authorization by the PMK (or spouse) to conduct the teacher and/or school principal questionnaire.

School Board consent rate is the rate of non-response that refers to the authorization given by the school board to conduct the teacher and/or school principal questionnaire.

Coverage is the extent to which the sample can adequately represent all elements of the target population.

10.3 Consents

For each child sampled, consent was requested from the parents, and then from the school boards to put the questions to the teachers and school principal.

The consent rate for the parents was very high: 94.8% nationally. Taken individually, there was no variable that presented very different consent rates as compared to the national rate. For the children whose parents gave their consent to contacting the schools, 96% of them were in educational institutions for which the school board had also given its agreement. The agreement of the school boards was established based on sending questionnaires to the teachers or to the principals.

The response rates by school boards differed slightly from one province to another and by type of school.

Province	School type 1 = Non-sectarian 2 = Sectarian	Weighted response rate
NL	1	98.7%
PE	1	100%
NS	1	91.4%
NS	2	100%
NB	1	98%
QC	1	97.6%
QC	2	89.1%
ON	1	93.8%
ON	2	94.6%
MB	1	95.5%
MB	2	100%
SK	1	97.3%
SK	2	92.9%
AB	1	94.6%
AB	2	96.7%
BC	1	92.6%
BC	2	86.4%
		95.9%

10.4 Non-response by teachers and school principals

The observation unit was a **respondent** if the teacher **OR** the school principal responded.

The following provides a summary of the responses of the teachers and school principals:

	Responses	Teachers	Principals
6,582 children for whom the teachers or the principal received a questionnaire	5,196 responses	2,940 single teachers	2,605 responses
			335 non-responses
		1,497 multi-teachers	1,359 responses
		No teacher response	138 non-responses
	1,386 non-responses		759 responses

The response rate varied between provinces, based on the size of the area of residence and the type of school (80% response for non-religious schools and 68% for religious schools).

Province	Response rate for teachers and principals
NL	88.8%
PE	92%
NS	84.3%
NB	87.1%
QC	78.9%
ON	78%
MB	77.1%
SK	76.6%
AB	70%
BC	70.3%
Canada	79%

Size of area of residence	Response rate for teachers and principals
rural region	82%
urban region < 30,000	82.1%
urban region 30,000 – 99,999	72.5%
urban region 100,000 – 499,999	80.3%
urban region more than 500,000	72.4%
Not stated	81.6%
Canada	79%

In cases where the teacher did not respond but the principal responded, or vice versa, the record for the child was considered to be a response. The data from the non-responding portion were then imputed so that the record would contain complete data (see section 11).

11.0 Imputation

As indicated in the previous section, a respondent was defined as being a child or youth for whom at least the teacher or the principal responded to the survey. Total non-response imputation was done when either portion was missing. Partial non-response for all the data relating to the child was not processed by imputation. Imputation is the process whereby missing items are "filled in" with plausible values.

For details about imputation, please refer to the main user guide.

11.1 Imputation of responses from school principals

Out of 5,196 questionnaires considered as respondents for the education portion, 4,437 (85%) had information on the teachers. Of these, 473 questionnaires (11%) did not contain information from the school principal. It was these 473 questionnaires which were subject to donor imputation.

To link donors to recipients, we tried to see if the responses of the teachers, for which a response from the principal was available, were similar to the responses from teachers without data from the principal.

The imputation classes were, at a minimum, created by province, type of school and size of area of residence. Using this strategy, there were few cases where the number of school principal questionnaires to be imputed was very high. Certain questionnaires had an unidentified area of residence. In these specific cases, we used the most frequent area of residence in the province and type of school.

In the imputation process, the minimal classes were created on the basis of province, type of school and size of area of residence. Other variables were also used: these dealt with the perceptions of teachers regarding the organization of the school, the support, team spirit, etc., and as well the children's attention in the classroom. These variables were different if the single or multiple teacher questionnaire was used. As well, a new school level indicator was created based on the child's age, to separate the primary and secondary levels. This flag was set to one if the child was aged 11 and over, otherwise it was set to 0. This variable was used in the imputation process.

The imputation process used the Impudon macro. In the first phase, a donor was assigned to the recipient who had the same characteristics for the variables: type of school, province, size of area of residence and child aged 11 and over or not.

In the process, we specified that we wanted similarity in the variables of teachers' perceptions between recipients and donors. All the variables of perception had an identical weight for the calculation of the distance, that is, we did not give preference to one variable over another in the calculation of the distance between two observations.

For each variable of similarity, we calculated the relative deviation between the value of the recipient and that of the donor. Then we calculated an overall distance for all the variables of perception, which is the weighted sum of the relative deviations and therefore a form of total distance based on the fields of similarity. The donor closest to the recipient in terms of that distance was selected.

We further questioned whether there was an agreement for child's age between donor and recipient.

This first step located 425 donors for 470 recipients. Out of 425 donors, 385 were used once, 35 twice, and 5 three times.

We performed a second imputation run to find a donor for the 3 questionnaires remaining to be imputed. For this run, we no longer required agreement for child's age. The same donor was used for these three records.

11.2 Imputation of the data from the teachers

Of the 5,196 responses to the child's questionnaire, 3,964 questionnaires (2,605+1,359) had responses from the teacher and the principal, and 759 questionnaires did not have a response from the teachers but did have a response from the principals. The concept was therefore to impute the response of the teachers for these 759 questionnaires.

The imputation classes were established by cross-classifying the **province, type of school, size of area of residence and age group (over the age of 11 or not)**. Here again, the imputation process used the Impudon macro.

Donor frequency	Number of donors
1	574
2	67
3	9
4	5

After the first run of the macro, we were able to link a donor to 755 of the 759 non-respondent teachers. There remained a need to find a donor for 4 recipients.

The second run was made by widening the constraint associated with the size of the area of residence; the exact agreement between size of area of residence of the donor and the recipient was no longer required. Proximity between the values of this variable for donor and recipient was enough to link a donor to a recipient. We then found a donor to impute 3 recipients. This last case to be imputed was the only one from a Catholic school; it was imputed using a neighbouring non-religious school.

12.0 Weighting and treatment of coverage and non-response

The aim of the weighting phase was to produce individual weightings for each respondent child. The sample of respondent children thus represented the population of children living in one of the 10 Canadian provinces, aged 6 to 15, and attending a public school. Each child was therefore assigned a final weighting, which took into account all the steps in the process.

The weightings produced were developed using the cross-sectional weight of the child in cycle 4 of the NLSCY. Only a **final cross-sectional weight was produced**.

12.1 Adjustment of the under-coverage

To adjust the under-coverage, a probability of being under-sampled is calculated for each eligible child. This probability is the result of logistic modelling based on province of residence or size of the area of residence and age of child.

The model is provided by:

$p = \text{Probability (unit was under-sampled)}$

and $\text{logit}(p) = 0.7979 + 0.6163 * (\text{Resides in NL}) + \dots + 0.3270 * (\text{Child aged 15})$.

Parameter	Estimate	Error	Chi-Square	Pr > ChiSq
Constant	0.7979	0.0505	249.4939	<.0001
<i>Reference = Québec</i>				
NL	0.6163	0.1858	11.0068	0.0009
PE	1.0762	0.3847	7.8259	0.0052
NS	0.3972	0.1403	8.0075	0.0047
NB	0.6405	0.1619	15.6586	<.0001
ON	-0.5118	0.0556	84.7562	<.0001
MB	0.6752	0.1274	28.0678	<.0001
SK	0.5093	0.1293	15.5189	<.0001
AB	0.1876	0.0795	5.5697	0.0183
BC	0.2067	0.0767	7.2711	0.0070
<i>Reference: urban area 30,000 – 99,000</i>				
Urban area 100,000 – 499,000	0.2180	0.0557	15.3115	<.0001
Urban area not stated	0.4307	0.1992	4.6728	0.0306
<i>Reference: children aged 9 and 12 years</i>				
AGE6	-0.3617	0.0737	24.1122	<.0001
AGE7	-0.6191	0.0695	79.4446	<.0001
AGE8	-0.4073	0.0696	34.2719	<.0001
AGE13	0.2310	0.0781	8.7442	0.0031
AGE14	0.2207	0.0765	8.3122	0.0039
AGE15	0.3270	0.0793	17.0115	<.0001

The model revealed an interesting phenomenon regarding under-coverage based on child's age: the greater the age of the child, the higher the probability of the unit being sampled. With regard to provinces, Ontario had a confirmed under-coverage. The model associated a probability of under-sampling with each eligible child. Five classes were established, each made up of children whose estimated probability was between the i quantile and the $i+1$ quantile. Here we used quantiles in the order of 20%, 40%, 60% and 80%.

On the model above, the quantiles for the estimated probability of being under-sampled were: 0.595(20%), 0.682(40%), 0.734(60%) and 0.8 (80%).

Given that the probability of not being under-sampled was, on average, less than that of being under-sampled, the class (p between 0-20%) should contain proportionally more under-sampled children.

Number of children under-sampled	P estimated between					Total
	80-100%	60-80%	40-60%	20-40%	0-20%	
No	370	456	550	753	1,009	3,138
Yes	1,845	1,591	1,416	1,250	1,136	7,238
Total	2,215	2,047	1,966	2,003	2,145	10,376
Average adjustment factor	1.21	1.31	1.39	1.6	1.89	1.5

For each under-sampled child, the weighting was adjusted within the class. The adjustment factors on average were between 1.2 to 1.9.

12.2 Adjustment of the parents' consent

The data were reduced to under-sampled children, then were processed using Knowledge Seeker to develop response classes that were considered to be homogenous. The variables used for the classification were:

- Adult smoking;
- Area of residence;
- Religion;
- Status of occupation of dwelling;
- Type of dwelling;
- Siblings;
- Involvement by the adult in sports activities with the child;
- Child's health;
- Adequate income level;
- Province;
- Sex;
- Child's activities with his friends;
- Born in Canada or not;
- Adult is a labour force participant or not;
- Adult depressed.

The variable of interest was the parents' consent.

Forty-four (44) homogenous response groups were established. We required a minimum of 30 observations per group with a response rate of at least 60% in general. We sometimes accepted lower response rates.

The highest weighted non-consent rate was 52.6%. This group included 39 children from the sample: 28 for whom the parents had given their consent and 11 for whom the parents had refused their consent. These were girls who had a brother or sister, who lived in an urban area of 500,000 residents or more, the reference adult did not smoke, and whose religion was Presbyterian, Baptist or Jehovah's Witness. The adjustment factors varied from 1 to 2.11.

12.3 Adjustment of the consent of the school boards

Out of 6,865 children whose parents gave their consent, 6,582 children were in educational institutions for which the school board had given its agreement.

The agreement of the school boards was established based on sending questionnaires to the teachers or principals.

The adjustment for non-response was made solely by province and by type of school. The adjustment factors were fairly low given the high response rate.

Province	Type of school 1 = Non-sectarian 2 = Sectarian	Number of questionnaires where parents gave consent	Weighted response rate	Adjustment factor
NL	1	445	98.7%	1.013
PE	1	263	100%	1.000
NS	1	536	91.4%	1.094
NS	2	1	100%	1.000
NB	1	483	98%	1.0204
QC	1	1,148	97.6%	1.0242
QC	2	68	89.1%	1.1198
ON	1	951	93.8%	1.0640
ON	2	475	94.6%	1.0563
MB	1	594	95.5%	1.0467
MB	2	1	100%	1.000
SK	1	483	97.3%	1.027
SK	2	132	92.9%	1.080
AB	1	507	94.6%	1.057
AB	2	171	96.7%	1.032
BC	1	581	92.6%	1.0811
BC	2	25	86.4%	1.146
		6,865	95.9%	

12.4 Adjustment for non-response by teachers or principals

Out of these 6,582 questionnaires sent out, 6,278 corresponded to a questionnaire sent to the teacher and 5,038 to a questionnaire sent to the school principal. Given the definition of a respondent unit (the response of the teacher or principal was recorded), 5,196 children were considered to be respondents, corresponding to a weighted response rate of 79%. The adjustment for non-response was based on a distinction between respondent and non-respondent units. The analysis was performed using observed variables that the two sub-populations of respondents and non-respondents had in common. The significant variables were: the province, size of area of residence, child's age, social support, parental status (biological or not), child's siblings, occupation status, child's health, that of the reference person, adequacy of income/family needs, and employment status during the previous year.

The strongest significance of the response function was by province. The Atlantic provinces were grouped together, as were Manitoba and Saskatchewan, and, Alberta and British Columbia.

For Atlantic Canada, the response rate was much lower if the child was over the age of 11 (multi-teacher questionnaire), especially if the child was over the age of 11 and the depression score was high;

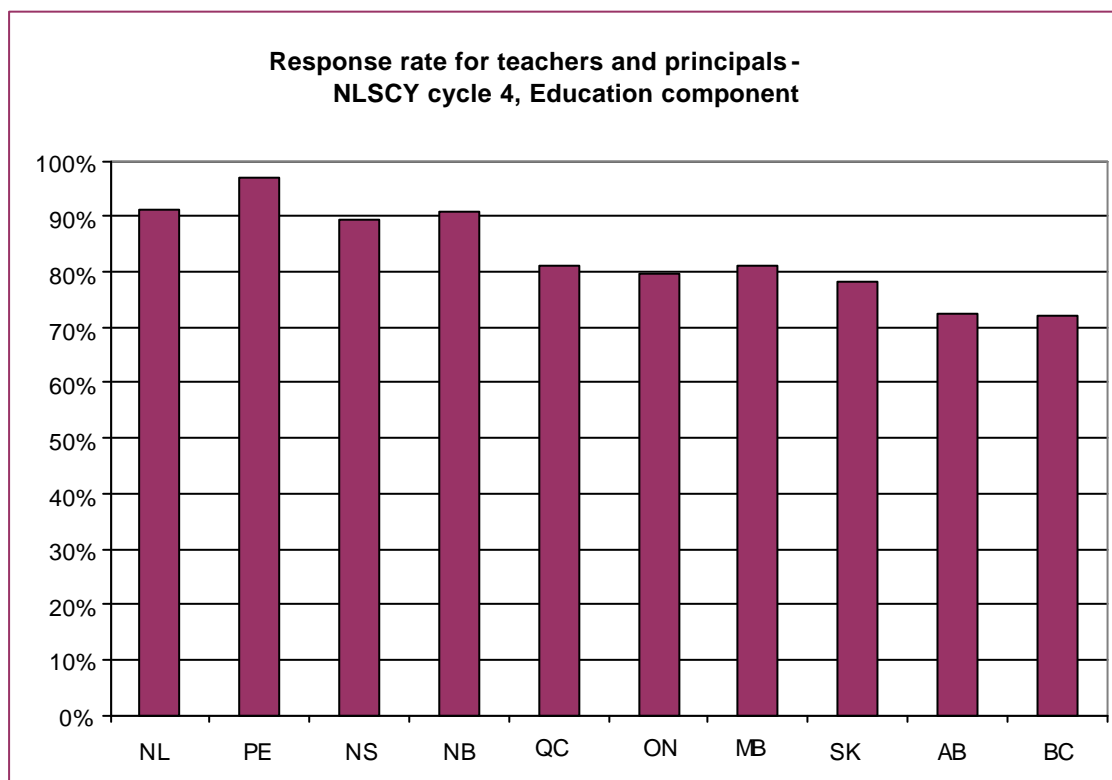
In Quebec, the significant variables for the response rate were: visiting the public library, parental status (biological, etc.), siblings, size of the area of residence, and emotional behaviour of the child. The lowest response rate was for children visiting the public library less than once or twice a year and having only one biological parent;

In Ontario, size of the area of residence influenced the response behaviour, while in the three largest cities, emotional behaviour of the child and siblings explained the response behaviour. In the other cities, response was characterized by marital status of the reference person, if he/she was born in Canada, not answering the questions on depression, and employment status.

In Manitoba and Saskatchewan, for those over the age of 11, emotional behaviour of the child was significant. For those under the age of 11, participating in sports or games with the adult influenced the response.

In Alberta and British Columbia the significant variables were: size of the area of residence, adult's health and employment status, marital status, and emotional behaviour of the child.

The average adjustment was 1.24. It was highest in the Western provinces and lowest in the Atlantic provinces.



12.5 Summary of adjustments

In summary, the four adjustments (under-coverage, consent of the school boards, parents' consent, response of teachers and principals) corrected on average the initial weights for the children by a factor of 1.99. This total adjustment varied by province; adjustments for under-coverage and for non-response by teachers or principals were the most significant.

Population	Sample size	Average adjustment			
		1 : under-coverage	2 : consent of school boards	3 : consent of parents	4 : response, teachers and principals
Total	5,196	1.44	1.06	1.05	1.24
from non-sectarian	4,636	1.41	1.06	1.04	1.24
from sectarian schools	560	1.63	1.06	1.06	1.30
NL	388	1.26	1.05	1.01	1.11
PE	242	1.22	1.06	1.00	1.11
NB	420	1.33	1.05	1.10	1.11
NS	413	1.25	1.06	1.02	1.11
QC	944	1.48	1.05	1.03	1.24
ON	1,054	1.76	1.05	1.06	1.28
MB	444	1.26	1.07	1.05	1.25
SK	449	1.30	1.06	1.04	1.25
AB	442	1.47	1.08	1.05	1.37
BC	400	1.44	1.08	1.09	1.47

12.6 Limitations

As stated before, the main goal in re-weighting the education file was to provide some respite for analysts having to deal with the reduced sample. This sub-sampling only affected the data collected for cycle 4 and so the focus was to ameliorate the potential for analysis at the cross-sectional level only. Through consultations, we identified a number of variables from the main component that were likely to be used in most analysis. Some of these variables were included in an iterative proportional fitting of the weights to maintain a certain degree of consistency between point estimates from the sub-sample and the full sample.

By including variables from the main component of this file it is hoped that users will be able to do most of their cross-sectional analysis without the main component. Should data users require the main file, they should note that their estimates may differ for variables not used for calibration, since any coverage differential from the main and sub-sampled files would not be accounted for by the adjusted education weight. Similarly, any longitudinal analysis will require the main file in order to obtain the longitudinal weight; the education information will only apply to a subset of the entire file, with no adjustment to account for the missing information.

Readers should note that, although every effort was made for iterative proportional fitting of the sub-sample to the characteristics of the full sample, certain biases may not be fully accounted for by the new weights.

13.0 Data quality and Coverage of the Main Survey

Information about coverage of the education portion of the files has been addressed in Chapter 10 of the present User Guide. For other details please refer to the Main User Guide.

14.0 Guidelines for Tabulation, Analysis and Release

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

14.1 Rounding Guidelines

In order that estimates for publication or other release derived from the NLSCY microdata file 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 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.
- d) Sums and differences of aggregates (or ratios) are to be derived from their corresponding unrounded components and then are to be rounded themselves to the nearest 100 units (or the nearest one decimal) using normal rounding.
- e) In instances where, due to technical or other limitations, a rounding technique other than normal rounding is used resulting in estimates to be published or otherwise released which differ from corresponding estimates published by Statistics Canada, users are urged to note the reason for such differences in the publication or release document(s).
- f) Under no circumstances are unrounded estimates to be published or otherwise released by users. Unrounded estimates imply greater precision than actually exists.

14.2 Sample Weighting Guidelines for Tabulation

Sample Design

The sample design used for the NLSCY was not self-weighting. When producing simple estimates, including the production of ordinary statistical tables, users must apply the proper survey weight. If proper weights are not used, the estimates derived from the microdata file cannot be considered to be representative of the survey population, and will not correspond to those produced by Statistics Canada. The weight assigned to each child reflects the number of children represented by a particular respondent.

For any analysis dealing with correlation analysis or any other statistics where a significance measure is required, it is recommended that a "sample" weight be used. This weight is obtained by multiplying the survey weight by the sample size and dividing this total by the total

estimated population. This produces a mean weight of 1 and a sum of weights equal to the sample size.

Benefit of Using an Adjusted or Rescaled Weight

For any analysis dealing with correlation or other statistic where a significance measure is required, it is recommended that “analytical” weights be used. These are the original weights rescaled to produce sample counts in lieu of population counts. The analytical weight is obtained by multiplying the demographic load by the sample size and dividing the product by the total estimated population. The result is a mean weight of 1 and a sum of weights equal to the sample size.

One of the benefits of using the adjusted weight is to circumvent an over estimation of the significance during estimation. While the estimate produced using this technique is likely underestimate the true measure of the standard error, it does provide a good initial indicator of significance when doing exploratory analysis. Users should use the recommended methods (e.g., using the bootstrap weight) to determine the significance of their findings before finalizing their report.

Cautionary note of using the adjusted weight.

The disadvantage to the adjusted weight is that the numerator is not weighted up to the target population; the Coefficient of Variance Tables, described in section 12 and presented in Appendix 3 of the Main User Guide, are no longer useful as a measure of data quality.

The survey methodologists recommend using the Bootstrap weights and the accompanying program to calculate more accurate estimates of the variance.

It should be noted, sub-domains resulting from the exclusion of partial non-response are not accounted for in this rescaled weight. This weight rescaling does not redistribute the demographic load of the units with missing information unless one assumes that it is truly random.

Software Differences

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

14.3 Definitions of Types of Estimates: Categorical vs. Quantitative

Unit of Analysis

The NLSCY file has been set up so that the child is the unit of analysis. The weights which appear on each record (DEDUW01C for the cross-sectional sample) is a weight corresponding to one child. It is therefore impossible to calculate the estimates with regard to parents and families based on the microdata file from the Education component of the NLSCY.

14.3.1 Categorical Estimates

Categorical estimates are estimates of the number, or percentage of the surveyed population possessing certain characteristics or falling into some defined category. An estimate of the number of persons possessing a certain characteristic may also be referred to as an estimate of an aggregate.

Examples of questions of the nominal type

Q: In your opinion, to what extent are the parent(s) or guardians involved in this child's education?

1. Very involved
2. Fairly involved
3. Not involved
4. I do not know/ no opinion

Q: What influence do these boards, associations or committees have on the policies or practices of your school?

1. Very large influence
2. Large influence
3. Limited influence
4. Little influence
5. No influence

14.3.2 Quantitative Estimates

Quantitative estimates are estimates of totals or of means, medians and other measures of central tendency of quantities based upon some or all of the members of the surveyed population. They also specifically involve estimates of the form $\frac{\hat{X}}{\hat{Y}}$ where

\hat{X} is an estimate of the surveyed population total quantity and \hat{Y} is an estimate of the number of people in the surveyed population contributing to that total quantity.

An example of a quantitative estimate is the average number of students in a classroom. The numerator is an estimate of the total number of students and the denominator is the number of classes (obtained from the number of responding teachers).

Example of quantitative questions

How many students are there in this student's classroom?

students

14.3.3 Tabulation of Categorical Estimates

Estimates of the number of children 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. These estimates may be cross-sectional or longitudinal.

Proportions and ratios of the form $\frac{\hat{X}}{\hat{Y}}$ are obtained by:

- a. summing the final weights of records having the characteristic of interest for the numerator (\hat{X});

- b. summing the final weights of records having the characteristic of interest for the denominator (\hat{Y}), then;
- c. dividing the numerator estimate by the denominator estimate.

14.3.4 Tabulation of Quantitative Estimates

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

For example, to obtain an estimate of the total number of students who present problems in speech, you must:

multiply the number of students in the class with speech problems by the final weight; then add this value for all the records indicating at least one student with a speech problem in the class.

In this student's class, how many students (including this student) present one or the other of the following chronic problems and has this problem been diagnosed or not by a professional? (Certain students may form part of more than one category.)

- a) Speech problem, auditory or visual problem, reduced mobility or other health problem which damages the student's learning?

students

The numerator \hat{X} is calculated in the same way as for a quantitative estimate and the denominator \hat{Y} is calculated in the same way as a nominal estimate. For example, to estimate the average number of children in a class who have a speech difficulty, you must:

- a) estimate the total number of students with a speech difficulty using the method described above;
- b) estimate the number of children in this category by adding the final weights of all the records corresponding to children who have at least one student in their class with a speech difficulty;
- c) Divide the obtained estimate in a) by the result of the calculation in b).

14.4 Guidelines for Statistical Analysis

Sample Design

The NLSCY 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.

Variance Estimates

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 variance estimates that are calculated are not adequate. Variances for simple estimates such as totals, proportions and ratios (for qualitative variables) are provided in the accompanying Sampling Variability Tables.

Rescaling or Adjusting the Weights

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

Example of Rescaling the Weights

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

- Select all respondents from the file with SEX = male (variable DMMCQ02).
- Calculate the AVERAGE weight for these records by summing the original person weights (DEDUW01C) from the microdata file for these records and then dividing by the number of records with SEX = male.
- For each of these records, 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's design are still not taken into account, the variances calculated in this way are likely to be under-estimated.

Calculation of Variance Estimates

The calculation of truly meaningful variance estimates requires detailed knowledge of the design of the survey. Such detail cannot be given in this microdata file because of confidentiality. Variances that take the complete sample design into account can be calculated for many statistics by Statistics Canada on a cost-recovery basis. For users who have access to microdata, better estimates of the variance can be produced for certain statistics by using the Bootstrap weights. (See section 15.9.)

14.5 Coefficient of Variation Release Guidelines

Release Guidelines

Before releasing and/or publishing any estimate from the NLSCY, users should first determine the quality level of the estimate. The quality levels are acceptable, marginal and unacceptable. As discussed in Chapter 10 in the Main guide, sampling and non-sampling errors both influence data quality. For the purposes of this document, however, estimate quality is based solely on the sampling error illustrated by the coefficient of variation, as shown in the table below.

First, the number of children who contribute to the calculation of the estimate should be determined. If this number is less than 30, the weighted estimate should be considered to be of unacceptable quality.

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

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

14.6 Quality Level Guidelines

Quality Level of Estimate	Guidelines
1. Acceptable	<p>Estimates have: a sample size of 30 or more, and low coefficients of variation in the range 0.0% to 16.5%.</p> <p>No warning is required.</p>
2. Marginal	<p>Estimates have: a sample size of 30 or more, and high coefficients of variation in the range 16.6% to 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 very high 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 the following warning should accompany the estimates:</p> <p>“The user is advised that...(specify the data)...do not meet Statistics Canada’s quality standards for this statistical program. Conclusions based on these data will be unreliable, and most likely invalid. These data and any consequent findings should not be published. If the user chooses to publish these data or findings, then this disclaimer must be published with the data.”</p>

15.0 Variance calculation

The NLSCY is a probabilistic survey. A sample has been selected to represent the target population. A given variability is inherent to the sample selection, which would be different if a different sample had been selected. This variability is known as the sampling error, as described in Chapter 10 in the Main guide. In addition, adjustments have been made to take into account the non-response. The measure of this variability is known as the variance. In this chapter, we will explain why it is important to calculate the variance and we will present different tools to do so.

15.1 Importance of the variance

The variance of an estimate is a good indicator of the quality of the estimate. A high variance estimate is considered non reliable. In order to quantify a large variance, a relative measure of the variability is used, namely the coefficient of variation (CV). The coefficient of variation is defined as the ratio of the square root of the variance over the estimate. The square root of the variance is also known as the standard deviation. The coefficient of variation, as opposed to the variance, allows the analyst to compare estimates of different magnitude on the same scale. As a result, it is possible to assess the quality of any estimates with the CV.

Also, the variance or the CV is required for statistical tests such as hypothesis tests, which determine if two estimates are statistically different. Consequently, variance or CV calculation is mandatory.

15.2 Variance and CV calculation

It is almost impossible to derive an exact formula to calculate the variance for the NLSCY, due to the complex sample design, non-response adjustments and the post-stratification. A very good way to approximate the true variance is to use the Bootstrap method. A set of 1,000 Bootstrap weights is available. Variance calculation using these 1,000 Bootstrap weights involves calculating the estimates with each of these 1,000 weights and then calculating the variance of these 1,000 estimates.

Two tools, all making use of the bootstrap weights, have been developed to help the users to calculate the variance and the CV for their estimates. These tools are:

- CV look-up tables, using a representative design effect, to get approximate CVs for some domains, by age cohort or by province.
- Macros to calculate the variance, using the Bootstrap weights.

The use of one or more of these tools depends on the type of analysis and the level of precision required.

15.2.1 CV look-up tables

In order to supply coefficients of variation which would be applicable to a wide variety of categorical estimates produced from this microdata file and which could be readily accessed by the user, a set of Approximate Sampling Variability Tables has been produced. These “look-up” tables can be found in the Research Data Centres, in Word format, under the filename of “TabVar_EDU_C4_E.doc”. They allow the user to obtain an approximate coefficient of variation based on the size of the estimate calculated from the survey data.

15.2.2 Computation of Coefficients of Variation

The coefficients of variation (CV) are derived using the variance formula for simple random sampling and incorporate a factor which reflects the multi-stage, clustered nature of the sample design. This factor, known as the design effect, was determined by first calculating design effects for a wide range of characteristics and then choosing from among these a conservative value to be used in the look-up tables which would then apply to the entire set of characteristics.

15.2.3 Sample Requirements

For the NLSCY, the sample was constructed taking account the following requirements.

- A sufficient sample was required in each of the 10 provinces to allow for the production of reliable estimates for all longitudinal children who were 0 to 11 years of age in Cycle 1.
- It was also necessary to have a large enough sample to produce estimates for Cycle 1 at the Canada level by seven key age groups or cohorts: 0 to 11 months, 1 year, 2 to 3 years, 4 to 5 years, 6 to 7 years, 8 to 9 years, and 10 to 11 years.
- In each province, a sufficient sample size was required for Cycle 2 to produce reliable estimates for all children who were 0 to 11 years of age in Cycle 1.

15.2.4 Design Effect, Sample Size, Population

The tables that follow show the design effects, sample sizes and population counts by province and age groups used to produce the Approximate Sampling Variability Tables. These tables reflect only the **cross-sectional** sample of the Education component.

Province	Design effect	Sample size	Population
Newfoundland and Labrador	2.1	388	67,135
Prince Edward Island	1.8	242	19,082
Nova Scotia	1.8	420	112,090
New Brunswick	2.2	413	90,501
Quebec	2.6	944	789,907
Ontario	2.9	1,054	1,431,105
Manitoba	2.2	444	147,872
Saskatchewan	2.3	449	143,866
Alberta	2.1	442	382,874
British Columbia	2.2	400	440,982
Atlantic	2.3	1,463	288,808
Others	2.7	2,398	2,661,994
Prairies	2.9	1,335	674,612
Total CANADA	3.7	5,196	3,625,414

Age	Design effect	Sample size	Population
Children 6-11 years	4.0	3,406	2,231,570
Children 12-15 years	2.9	1,790	1,393,844
Total CANADA	3.7	5,196	3,625,414

Age	Design effect	Sample size	Population
Children 6-7 years	2.9	1,322	700,751
Children 8-9 years	3.8	1,048	790,875
Children 10-11 years	3.7	1,036	749,612
Children 12-13 years	2.8	915	734,200
Children 14-15 years	3.0	875	649,976
Total CANADA	3.9	5,196	3,625,414

15.2.5 Approximate Sampling Variability Tables

All coefficients of variation in the Approximate Sampling Variability Tables are approximate and, therefore, unofficial. The use of actual variance estimates would likely result in estimates with lower variances; for example, estimates listed as “unacceptable” in the Approximate Sampling Variability Tables could move up to the “marginal” category.

Remember: If the number of observations on which an estimate is based is less than 30, the weighted estimate should be classified as “unacceptable” regardless of the value of the coefficient of variation for this estimate. This is because the formulas used for estimating the variance do not hold true for small sample sizes.

15.3 How to Use the CV Tables For Categorical Estimates

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

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

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

Rule 2: Estimates of Proportions or Percentages Possessing a Characteristic

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

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

When the proportion or percentage is based upon a subset of the total population (for example, those who belong to a specific age category or sex in a given province), reference should be made to the proportion or percentage (across the top of the table) and to the numerator of the proportion or percentage (down the left side of the table). The intersection of the appropriate row and column gives the coefficient of variation.

Rule 3: Estimates of Differences Between Aggregates or Percentages

The standard error of a difference between two estimates is approximately equal to the square root of the sum of squares of each standard error considered separately.

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

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

where \hat{X}_1 is estimate 1, \hat{X}_2 is estimate 2, a_1 and a_2 are the coefficients of variation of \hat{X}_1 and \hat{X}_2 respectively. The coefficient of variation of \hat{d} is given by $s_{\hat{d}}/\hat{d}$. This formula is accurate for the difference between separate and uncorrelated characteristics, but is only approximate otherwise.

Rule 4: Estimates of Ratios

Where the numerator is not a subset of the denominator, the standard deviation of the ratio of the estimates is approximately equal to the square root of the sum of squares of each coefficient of variation considered separately multiplied by the ratio itself.

The standard error of ratio $\hat{R} = \hat{X}_1/\hat{X}_2$ is therefore:

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

where a_1 and a_2 are coefficients of variation of \hat{X}_1 (number of girls demonstrating learning problems) and \hat{X}_2 (number of boys demonstrating learning problems) respectively. The coefficient of variation of \hat{R} is given by $s_{\hat{R}}/\hat{R}$. The formula will tend to overstate the error, if \hat{X}_1 and \hat{X}_2 are positively correlated and understate the error if \hat{X}_1 and \hat{X}_2 are negatively correlated.

Rule 5: Estimates of Differences of Ratios

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

Warning Note on Confidence Intervals

Release guidelines applying to estimates also apply to confidence intervals. For example, if the estimate is “marginal”, then the confidence interval is marginal and should be accompanied by a warning note to caution subsequent users about high levels of error.

15.4 Examples of use of the CV tables and of the coding for the nominal estimates

The examples provided below are intended to assist users in applying the rules we have just presented.

Example 1: Estimates of the number of persons possessing a specific characteristic (aggregates)

Assume that, based on the data from the Education component of the NLSCY, it is estimated that 92,342 children aged 12 and over have repeated a grade. How can the user determine the CV of this estimate?

1. Turn to the CV table for children aged 12 to 15.
2. The estimated aggregate (92,342) is not found in the column to the left, Percentage numerator; you must therefore use the number closest to it, i.e. 90,000.
3. The CV for an estimated aggregate is the first record, other than the asterisks, on this line, i.e. 14.4 %.
4. The approximate CV of the number of children whose reference person responded that the child has had an average degree of overall success is therefore 14.4 %. The result is acceptable and no warning is required to release the estimate, given that the CV of the estimate is within the 0.0 % to 16.5 % range.

Example 2: Estimates of the proportions or percentages of persons possessing a specific characteristic (aggregates)

Assume that, based on the data from the Education component of the NLSCY, it is estimated that 32.2% of the 92,342 children aged 12 and over, who have repeated a grade, have a speech problem. How can the user determine the CV of this estimate?

1. Turn to the CV table for children aged 12 to 15.
2. Because the estimate is a percentage based on a subset of the total population, you must use both the percentage (32.2 %) and the numerator portion of the percentage (29,734) to determine the CV.
3. The numerator, 29,734 does not appear in the column to the left, Percentage numerator; you must therefore use the number closest to it, i.e. 30,000. As well, the estimate of the percentage does not appear in the column headings and therefore you must use the number closest to it, i.e., 30,000.
4. The number found at the intersection of the line and column used, i.e. 22.5%, is the CV to use.
5. The approximate CV is estimated at 22.5 %. You must therefore provide a warning when releasing the result, i.e. that 32.2% of the children aged 12 and over and having repeated a grade, had speech problems. This estimate is “marginal”, given that the CV is in the 16.5 % to 33.3% range.

Example 3: Estimates of the differences between the aggregates or the percentages

Let us assume that, based on the data from the Education component of the NLSCY, we estimate that 60.5% of the children aged 12 and over, who have repeated a year are boys and 39.5% of children aged 12 years and over having repeated a grade are girls. How can the user determine the CV of the difference between these two estimates?

1. Using the CV table for children aged 12 to 15, in the same manner as in example 2, you establish a CV of the estimate of 10.9 % for girls and for boys.
2. The standard deviation of the estimate of the difference is therefore given by rule 3, which is,

$$s_d = \sqrt{(\hat{x}_1 a_1)^2 + (\hat{x}_2 a_2)^2}$$

$$\text{or, } ((0.605 * 0.109)^2 + (0.395 * 0.193)^2)^{1/2} = 10.1\%$$

Example 4: Ratio estimates

Now suppose that a user wishes to compare the number of girls aged 12 to 15 years having repeated a grade to the number of boys aged 12 to 15 having repeated a year. The user is interested in comparing these estimates as a ratio. How is the CV of this estimate calculated?

1. First of all, this is a ratio estimate, where the numerator of the estimate is the number of girls in the comparison and the denominator is the number of boys.
2. Turn to the CV table for the 12 to 15 age group.
3. The numerator of this ratio estimate is 36,475. The closest number to this is 35,000. The CV for this estimate is determined by finding the first record, other than asterisks, on this line, i.e. 24.3 %.
4. The denominator of this ratio estimate is 55,867. The closest number to this is 55,000. The CV for this estimate is determined by finding the first record, other than asterisks, on this line, i.e. 19.4 %.
5. The standard deviation of the ratio estimate is therefore given by rule 4, which is,

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

$$\text{or, } (36475/55867) * (0.194^2 + 0.243^2)^{1/2} = 0.203$$

The ratio of girls aged 12 to 15 years having repeated a grade to boys aged 12 to 15 years having repeated a grade is 36475 to 55867, or 0.65.

The CV of this estimate is 0.203/0.65 = 31.2% that is, “marginal”. This estimate may thus be released with a caution, given that the CV is in the 16.5% to 33.3% range.

15.5 How to use the CV tables to obtain confidence limits

Although coefficients of variation are widely used, a more intuitively meaningful measure of sampling error is the confidence interval of an estimate. A confidence interval constitutes a

statement on the level of confidence that the true value for the population lies within a specified range of values. For example a 95% confidence interval can be described as follows:

If sampling of the population is repeated indefinitely, each sample leading to a new confidence interval for an estimate, then in 95% of the samples the interval will cover the true population value.

Using the standard error of an estimate, confidence intervals for estimates may be obtained under the assumption that under repeated sampling of the population, the various estimates obtained for a population characteristic are normally distributed about the true population value. Under this assumption, the chances are about 68 out of 100 that the difference between a sample estimate and the true population value would be less than one standard error, about 95 out of 100 that the difference would be less than two standard errors, and about 99 out of 100 that the differences would be less than three standard errors. These different degrees of confidence are referred to as the confidence levels.

Confidence intervals for an estimate are generally expressed as two numbers, one below the estimate and one above the estimate, as where k is determined depending upon the level of confidence desired and the sampling error of the estimate.

Confidence intervals for an estimate can be calculated directly from the Approximate Sampling Variability Tables by first determining from the appropriate table the coefficient of variation of the estimate and then using the following formula to convert to a confidence interval (CI):

$$IC_X = \left[\hat{X} - t\hat{X}s_{\hat{X}}; \hat{X} + t\hat{X}s_{\hat{X}} \right]$$

where \hat{X} is the determined coefficient of variation, and

$t = 1$ if a 68% confidence interval is desired

$t = 1.6$ if a 90% confidence interval is desired

$t = 2$ if a 95% confidence interval is desired

$t = 3$ if a 99% confidence interval is desired.

Note Regarding Release Guidelines

Release guidelines applying to estimates also apply to confidence intervals. For example, if the estimate is “marginal”, then the confidence interval is marginal and should be accompanied by a warning note to caution subsequent users about high levels of error.

Example of Using the CV Tables to Obtain Confidence Limits

This is how a confidence interval (CI) of 95 % for the estimated proportion of children aged 12 to 15 years having repeated a year is calculated.

Estimate of $X = 4.4$ %, $t = 2$.

Estimate of alpha of $X = 14.8$ % is the coefficient of variation of this estimate as taken from the tables.

$$CI_X = \{0.44 - (2)(0.044)(0.148); 0.44 + (2)(0.044)(0.148)\}$$

$$CI_X = \{0.44 - 0.013; 0.44 + 0.013\}$$

$$CI_X = \{0.031; 0.057\}$$

We can state, with a probability of 95 %, that between 3.1 % and 5.7 % of children aged 12 years and older have repeated a year.

15.6 How to Use the CV Tables to Do a T-test

Hypothesis Testing

Standard errors may also be used to perform hypothesis testing, a procedure for distinguishing between population parameters using sample estimates. The sample estimates can be numbers, averages, percentages, ratios, etc. Tests may be performed at various levels of significance, where a level of significance is the probability of concluding that the characteristics are different when, in fact, they are identical.

Let \hat{X}_1 and \hat{X}_2 be sample estimates for two characteristics of interest. Let the standard error on the difference $\hat{X}_1 - \hat{X}_2$ be s_d .

If $t = \frac{\hat{X}_1 - \hat{X}_2}{s_d}$ is between -2 and 2, then no conclusion about the difference between the

characteristics is justified at the 5% level of significance. If however, this ratio is smaller than -2 or larger than +2, the observed difference is significant at the 0.05 level. That is to say that the characteristics are significantly different.

Example of Using CV Tables to do a T-Test

Suppose we wish to test, at 5% level of confidence, the hypothesis that there is no difference between the proportion of girls and boys repeating a grade among children aged 12 to 15 years. From example 3, the standard error of the difference between these two estimates was found to be = 10.1%.

Hence,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{s_d} = \frac{0.605 - 0.395}{0.101} = 2.08$$

Since $t = 2.08$ is outside -2 and 2, we can conclude at the 0.05 level of confidence that there is a difference between the proportion of boys among children aged 12 to 15 years having repeated a grade and the proportion of girls among children aged 12 to 15 years having repeated a grade.

15.7 Coefficients of Variation for Quantitative Estimates

Quantitative Estimates

For quantitative estimates, special tables would have to be produced to determine their sampling error. Since most of the variables for the NLSCY are categorical in nature, this has not been done.

As a general rule, however, the coefficient of variation of a quantitative total will be larger than the coefficient of variation of the corresponding category estimate. If the corresponding category estimate is not releasable, the quantitative estimate will not be either. For example, the coefficient of variation of the number of students 12 to 15 years would be greater than the coefficient of variation of the corresponding proportion of students 12 to 15 years. Hence if the coefficient of variation of the proportion is not releasable, then the coefficient of variation of the corresponding quantitative estimate will also not be releasable.

Pseudo Replication

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

15.8 Release Cut-offs for the NLSCY

In the tables that follow, cut-off numbers are given for NLSCY estimates in order for them to be of “acceptable”, “marginal” or “unacceptable” quality. Users are encouraged to use these cut-offs when publishing data from the NLSCY. First a table is given to show the cut-offs at the provincial, regional and Canada level. Then a table is given to show the cut-offs for the various age cohorts.

For example, an estimate for Nova Scotia of 5,000 would fall into the “marginal” range. This would mean that the estimate should be flagged and a note of caution would be attached for subsequent users about the high level of error associated with the estimate.

**Geographical Release Cut-Offs
Cross-Sectional Sample**

Province	Acceptable – Estimates at or above:	Marginal - Estimates between:	Unacceptable – Estimates at or below :
Newfoundland and Labrador	11,000	3,500 – 12,000	3,500
Prince Edward Island	4,500	1,500 – 4,500	1,500
Nova Scotia	15,500	4,500 – 1,5500	4,500
New Brunswick	15,000	4,500 – 15,000	4,500
Quebec	73,500	19,500 – 73,500	19,500
Ontario	132,000	35,000 – 132,000	35,000
Manitoba	23,500	6,500 – 23,500	6,500
Saskatchewan	23,000	6,500 – 23,000	6,500
Alberta	58,000	16,500 – 58,000	16,500
British Columbia	74,500	21,000 – 74,500	21,000
Atlantic Provinces	16,000	4,500 – 16,000	4,500
Other	107,500	27,500 – 107,500	27,500
Prairies	50,000	13,000 – 50,000	13,000
Canada	91,500	23,000 – 91,500	23,000

**Release Cut-Offs by Age Group
Cross-Sectional Sample**

Age Group	Acceptable – Estimates at or above:	Marginal - Estimates between:	Unacceptable – Estimates at or below :
Children aged 6 – 11 years	92,000	23,500 – 92,000	23,500
Children aged 12 – 15 years	79,000	20,500 – 79,000	20,500
Children aged 6 and 7 years	49,500	13,000 – 49,500	13,000
Children aged 8 and 9 years	83,000	22,500 – 83,000	22,500
Children aged 10 and 11 years	84,500	23,000 – 84,500	23,000
Children aged 12 and 13 years	71,000	19,000 – 71,000	19,000
Children aged 14 and 15 years	76,000	20,500 – 76,000	20,500

15.9 SAS and SPSS macros to calculate the variance, using the Bootstrap weights.

SAS and SPSS macros have been developed to calculate the variance using the Bootstrap weights. From the methods available, this method is the best approach to obtain a very good approximation of the true variance. With this method, it is possible to calculate the variance of any estimates, for any domains. The variance calculated using this method takes into account the sample design and also, the specificities of the variable of interest. Finally, as opposed to the other methods, the user is not restricted to pre-defined domains.

This method has many advantages but requires more work from the researcher. Variance calculation using these macros is more time consuming than the other two methods presented earlier. The user must first get familiar with the macros before using them. However, these macros have been developed in such a way that they are easy to use.

The researcher must have access to the macros, to the data files and to the Bootstrap weight files. Access to these tools is possible in the Statistics Canada Data Research Centers (RDC). Also, detailed documentation on how to use these SAS or SPSS macros is available in the RDCs.

Despite the time required to run these macros, it is strongly recommended to use this method to calculate the variance of any estimates which must be published. This method provides a more precise and accurate measure of the true variance. The Bootstrap weight file has 1,000 variables (BSW1 to BSW1000) corresponding to 1,000 bootstrap weights and child identification variables.

16.0 Direct Assessment

Research on early childhood development plays a significant role in the formulation of policy for young children. Adding early developmental assessment tools to the NLSCY will help to enhance the knowledge about developmental processes in early childhood and provide relevant data on which to base policy directions for this stage.

Choices about the assessment tools to be included in the NLSCY for Cycle 4 (collection in 2000/01) were made on the basis of an extended literature review, development of a research framework on child development and learning, consultations with many experts in Canada and internationally, review of material on many different possible instruments and field testing of the most likely possibilities. The instruments selected for consideration were also reviewed using a number of criteria. The criteria included reliability and validity of the instrument, coverage of domains in the research framework, ability of the instrument to indicate normal development and developmental delays, the ease of administration by lay interviewers and the availability of the instrument in English or French (or ease of translation to French or English). The final decision was strongly influenced by key experts who had had a history of providing advice to the NLSCY Team.

Due to the unique nature of this file, variables were included from the Main file to facilitate analysis. The following variables were included for each respondent where applicable: PPVT-R score (cycles 1 to 4), math score (cycles 1 to 4) and reading score (cycles 2 and 3). These measures are described in detail in the Main User Guide.

17.0 Analytic Issues

This chapter provides users with an overview of the various analytic issues one should consider when analyzing the Education component of NLSCY data. Some of the points mentioned in this chapter have been explored in greater detail in other chapters. The idea is to assemble in a single chapter all the important issues that might affect an analysis of the data.

17.1 How a complex sample design affects analysis

As described in the chapter 5 on the sample and the chapter 12 on weighting, the children in the NLSCY sample were selected using a complex sample design in order to meet clients' needs and address certain operational constraints. To make effective use of the data and produce valid estimates, it is important to use sample weights, either longitudinal or cross-sectional depending on the type of analysis.

17.2 Unit of analysis

In the NLSCY, the unit of analysis is always the child. Some household data were collected, but no estimates can be produced for the household level. All estimates must be for the child level.

17.3 Type of analysis: cross-sectional, longitudinal or repeated

With the sample reduction for the education component of the survey, analysts will have to contend with the effects of a smaller sample on the type of analysis than can be performed. Clearly with the current structure of this release, we have favoured the cross-sectional analysis of this data. Having said this, analysts are not limited to only that type of analysis.

With the NLSCY, users have their choice of different types of analysis: longitudinal analysis, cross-sectional analysis or even treating the NLSCY data as if they came from a repeated survey. Each type involves different issues.

17.3.1 Longitudinal analysis

In any longitudinal analysis, the population represented when using the longitudinal weights is always the population at the time of the child's initial selection. With the survey now in its fourth Cycle, the analyst must deal with the problems of sample attrition and lack of coverage due to total non-response and the problem of partial non-response. In addition, since the children selected in Cycle 1 have now responded four times, a conditioning bias may have developed over time. All these problems are described later in this chapter.

In addition, in Cycle 4, for the first time, two sets of longitudinal weights are available. The first set comprises longitudinal children who responded in Cycle 4, whether or not they responded in all previous cycles. The second set of longitudinal weights, computed for children introduced in Cycle 1, applies to children who have responded in every cycle.

17.3.2 Cross-sectional analysis

In any cross-sectional analysis, the population represented when using the cross-sectional weights is always the population at the time of collection of the current cycle

(Cycle 4). In addition to the above-mentioned problems affecting longitudinal analysis, a further problem must be taken into consideration in cross-sectional analysis: the coverage of the current population. As described in Chapter 5 of the Main guide, the cross-sectional sample is composed of children selected at different points in time. For example, children aged 6 to 17 in 2000 were selected in 1994. The sample has not been replenished since. Consequently, people who have immigrated to Canada since 1994 have no chance of being included in the sample. The impact of this situation is described in the chapter on data quality. Moreover the sub-sampling done for the education component (described in Chapter 5 of this guide) will have an effect on the precision of the estimates, especially for sub-domains, as the remaining sample size has been reduced.

17.3.3 Repeated analysis

A repeated analysis is an analysis in which a single target population is studied at different points in time. For example, with the NLSCY, one-year-olds can be studied in 1994/95, 1996/97, 1998/99 and 2000/01.

With the Education component, this type of analysis will be complicated as re-weighting was not performed in previous cycles. Uneven treatment on non-response will impact any analysis in this case. Users who wish to do this analysis would likely forego the adjusted weight, and revert to the Main weight by linking the subset to the entire sample. Depending on the domain being analyzed, certain issues can arise in either longitudinal or cross-sectional analysis. Other issues should also be considered. First, for a particular age group, the number of children in the sample can vary substantially from one cycle to another. This is especially true for children aged 1 to 5 years.

In addition, since the Education component applies only to the longitudinal portion of the sample in Cycle 4, analysts should realize that the samples are dependent from one cycle to the next. Children for whom education data are collected were in the sample in the previous cycle. Even when looking at separate age cohorts, many respondents may in fact be siblings of respondents in the next cycle.

A number of other issues are also discussed in the main guide.

17.4 Partial and total non-response

Like any other survey, the NLSCY is subject to non-response. There are two main types of non-response: total and partial. By definition, a respondent child has an adult component and at least one child/youth component completed for that household (it may be their own or a sibling's component).

17.4.1 Total non-response

Total non-response is the complete lack of data for a selected and eligible child due to factors such as refusal to take part in the survey or inability to trace the child. Total non-response is taken into account and corrected within the sampling weights. This process is described in detail in the chapter on weighting. The greater the amount of total non-response, the more difficult it is to adjust the weight of responding units to compensate for the non-response. Moreover, from a longitudinal standpoint, with the first set of longitudinal weights described above, a child who responded in Cycle 4 may not have responded in a previous cycle. The analyst must account for this in their analysis.

17.4.2 Partial non-response

For a child to be considered a respondent in Cycle 4, the child/youth component or the Person Most Knowledgeable (PMK) component must be completed. This means that a child can be deemed a respondent even if some questions or even entire components are not answered. For a given child, there are many sources of information. Depending on the child's age, the information may come from one or more of the following sources: themselves, a parent, a teacher, the school principal. Hence, even if one of these sources failed to provide the desired information, the child may still be considered a respondent.

In particular, the income variables are imputed. Most other variables are set to "don't know" or "not stated". The analyst must therefore determine the extent of non-response in the variables of interest before proceeding with the analysis.

17.5 Data processing

The NLSCY contains a huge amount of information. Although every effort is made to ensure that the data is of high quality with online editing while the data is being collected, if each variable were verified once captured, it would be impossible to make the data available within a reasonable length of time. As a result, data processing focuses on the survey's key variables, and errors may be left in the data. However, efforts are made to keep errors to a minimum.

17.6 Computing the variance with certain software applications

SAS and SPSS, software applications commonly used by analysts, are able to compute point estimates correctly using sampling weights. With the exception of some SAS procedures, these applications are not designed to take into account the sample design when calculating the variance. These applications do not take into account the estimation complexities of the NLSCY (complex sample design, non-response adjustment and post-stratification) in calculating the variance. As a result, they tend to underestimate the variance, sometimes substantially. Consequently, the analyst is strongly encouraged to use the Bootstrap weights for the variance estimation.

Some applications can use the Bootstrap weights. SUDAAN and WesVar take into account the sample design in calculating the variance, using the Balanced Repeated Replication (BRR) method. The creation of BRR weights differ in theory from the creation of Bootstrap weights, but the variance estimator is the same. As a result, the NLSCY Bootstrap weights can be used with these applications.

There is a stand alone version of SUDAAN, as well as an integrated version with SAS. The latter gives the flexibility to use the SUDAAN procedures within SAS. With its 9 procedures, SUDAAN can produce estimates of means, ratios and totals, independence tests, linear, log-linear and logistic regressions and survival tests. SUDAAN can read SAS and SPSS files, as well as other common types of files.

Wesvar uses a "point and click" approach, which makes it easy to learn. The types of analysis are more limited than SUDAAN, but WesVar still allows the estimates of means, ratios and totals, independence tests, as well as linear and logistic regressions. WesVar can read SAS, SPSS and other common type of files, but the application requires an additional step to create a WesVar file before processing with the analysis.

Bootstrap weights can be used with other applications which offer the required programming environment and the desired analytical tools. SAS and SPSS macros have been developed by NLSCY to use the Bootstrap weights to produce variance estimates based on the sample design.

18.0 Appendix I - Record Layouts

This appendix describes the layout of the files. It presents for each variable its position within the file, its name, its length and if it is an alpha or numeric value.

Example:

Variable START	Variable name	Variable length	Variable characteristics
@01213	DEDUW01C	12.4	Numeric value with four decimals and of a maximum of twelve bites including the decimal symbol.
@00840	DMMCQ02	\$1.	The symbol: \$ indicates an alpha value of a maximum of 1 bite.

18.1 Education File

@00001	DEDUTQ01	1.	@00041	DEDUSQ39	1.
@00002	DEDUTQ02	1.	@00042	DEDUSQ40	1.
@00003	DEDUTQ03	1.	@00043	DEDUSQ41	1.
@00004	DEDUTQ04	1.	@00044	DEDUSQ42	1.
@00005	DEDUTQ05	1.	@00045	DEDUTQ43	1.
@00006	DEDUTQ06	1.	@00046	DEDUTQ44	1.
@00007	DEDUTQ07	1.	@00047	DEDUSQ45	2.
@00008	DEDUTQ08	1.	@00049	DEDUSQ46	2.
@00009	DEDUTQ09	1.	@00051	DEDUSQ47	2.
@00010	DEDUTQ10	1.	@00053	DEDUSQ48	2.
@00011	DEDUTQ11	1.	@00055	DEDUSQ49	2.
@00012	DEDUMQ13	1.	@00057	DEDUSQ50	2.
@00013	DEDUTQ12	1.	@00059	DEDUSQ51	2.
@00014	DEDUSQ14	2.	@00061	DEDUSQ52	2.
@00016	DEDUMQ15	2.	@00063	DEDUTQ54	1.
@00018	DEDUSQ16	1.	@00064	DEDUTQ55	1.
@00019	DEDUSQ17	1.	@00065	DEDUTQ56	1.
@00020	DEDUSQ18	1.	@00066	DEDUTQ57	1.
@00021	DEDUSQ19	1.	@00067	DEDUTQ58	1.
@00022	DEDUSQ20	1.	@00068	DEDUTQ59	1.
@00023	DEDUSQ21	1.	@00069	DEDUTQ60	1.
@00024	DEDUSQ22	1.	@00070	DEDUTQ53	1.
@00025	DEDUSQ23	1.	@00071	DEDUTQ61	1.
@00026	DEDUSQ24	1.	@00072	DEDUTQ62	1.
@00027	DEDUSQ25	1.	@00073	DEDUTQ63	1.
@00028	DEDUSQ26	1.	@00074	DEDUTQ64	1.
@00029	DEDUSQ27	1.	@00075	DEDUTQ65	1.
@00030	DEDUSQ28	1.	@00076	DEDUTQ66	1.
@00031	DEDUSQ29	1.	@00077	DEDUTQ67	1.
@00032	DEDUSQ30	1.	@00078	DEDUTQ68	1.
@00033	DEDUSQ31	1.	@00079	DEDUTQ69	1.
@00034	DEDUSQ32	1.	@00080	DEDUSQ70	1.
@00035	DEDUSQ33	1.	@00081	DEDUTQ71	1.
@00036	DEDUSQ34	1.	@00082	DEDUMQ72	2.
@00037	DEDUSQ35	1.	@00084	DEDUTQ73	2.
@00038	DEDUSQ36	1.	@00086	DEDUTQ74	1.
@00039	DEDUSQ37	1.	@00087	DEDUTQ75	1.
@00040	DEDUSQ38	1.	@00088	DEDUTQ76	1.

@00089	DEDUTQ77	2.	@00184	DEDUT134	2.
@00091	DEDUMQ78	1.	@00186	DEDUT135	2.
@00092	DEDUMQ79	1.	@00188	DEDUT136	2.
@00093	DEDUMQ80	1.	@00190	DEDUT137	2.
@00094	DEDUMQ81	1.	@00192	DEDUT138	2.
@00095	DEDUMQ82	1.	@00194	DEDUT139	2.
@00096	DEDUMQ83	1.	@00196	DEDUT140	2.
@00097	DEDUMQ84	2.	@00198	DEDUS141	1.
@00099	DEDUTQ86	1.	@00199	DEDUS142	1.
@00100	DEDUSQ87	1.	@00200	DEDUS143	1.
@00101	DEDUSQ88	1.	@00201	DEDUS144	1.
@00102	DEDUSQ89	1.	@00202	DEDUS145	1.
@00103	DEDUSQ90	1.	@00203	DEDUS146	1.
@00104	DEDUSQ91	1.	@00204	DEDUS147	1.
@00105	DEDUSQ92	1.	@00205	DEDUS148	1.
@00106	DEDUSQ93	1.	@00206	DEDUS149	1.
@00107	DEDUSQ94	1.	@00207	DEDUS150	1.
@00108	DEDUSQ95	2.	@00208	DEDUS151	1.
@00110	DEDUSQ96	2.	@00209	DEDUS152	1.
@00112	DEDUSQ97	2.	@00210	DEDUS153	1.
@00114	DEDUSQ98	2.	@00211	DEDUS154	1.
@00116	DEDUSQ99	2.	@00212	DEDUS155	1.
@00118	DEDUT100	2.	@00213	DEDUS156	1.
@00120	DEDUMI01	2.	@00214	DEDUS157	1.
@00122	DEDUMI02	1.	@00215	DEDUS158	1.
@00123	DEDUMI03	1.	@00216	DEDUS159	1.
@00124	DEDUMI04	2.	@00217	DEDUS160	1.
@00126	DEDUMI05	2.	@00218	DEDUS161	1.
@00128	DEDUMI06	2.	@00219	DEDUS162	1.
@00130	DEDUMI07	2.	@00220	DEDUS163	1.
@00132	DEDUMI08	2.	@00221	DEDUS164	1.
@00134	DEDUMI09	2.	@00222	DEDUS165	1.
@00136	DEDUMI10	2.	@00223	DEDUS166	1.
@00138	DEDUMI11	2.	@00224	DEDUS167	1.
@00140	DEDUS112	2.	@00225	DEDUS168	1.
@00142	DEDUS113	2.	@00226	DEDUS169	1.
@00144	DEDUS114	2.	@00227	DEDUS170	1.
@00146	DEDUT115	2.	@00228	DEDUS171	1.
@00148	DEDUT116	2.	@00229	DEDUS172	1.
@00150	DEDUT117	2.	@00230	DEDUS173	1.
@00152	DEDUT118	2.	@00231	DEDUS174	1.
@00154	DEDUT119	2.	@00232	DEDUS175	1.
@00156	DEDUT120	2.	@00233	DEDUS176	1.
@00158	DEDUT121	2.	@00234	DEDUS177	1.
@00160	DEDUT122	2.	@00235	DEDUS178	1.
@00162	DEDUS123	2.	@00236	DEDUS179	1.
@00164	DEDUS124	2.	@00237	DEDUS180	1.
@00166	DEDUT125	2.	@00238	DEDUS181	1.
@00168	DEDUT126	2.	@00239	DEDUS182	1.
@00170	DEDUS127	2.	@00240	DEDUS183	1.
@00172	DEDUT128	2.	@00241	DEDUT184	2.
@00174	DEDUS129	2.	@00243	DEDUT185	2.
@00176	DEDUS130	2.	@00245	DEDUT187	2.
@00178	DEDUS131	2.	@00247	DEDUT188	2.
@00180	DEDUT132	2.	@00249	DEDUT189	1.
@00182	DEDUT133	2.	@00250	DEDUMI90	1.

@00251	DEDUM191	1.	@00341	DEDUM248	2.
@00252	DEDUM192	1.	@00343	DEDUM249	2.
@00253	DEDUM193	1.	@00345	DEDUM250	2.
@00254	DEDUM194	1.	@00347	DEDUS251	1.
@00255	DEDUM195	1.	@00348	DEDUS252	2.
@00256	DEDUM196	1.	@00350	DEDUS253	2.
@00257	DEDUM197	2.	@00352	DEDUS254	2.
@00259	DEDUT198	1.	@00354	DEDUS255	2.
@00260	DEDUS200	1.	@00356	DEDUS256	2.
@00261	DEDUS201	1.	@00358	DEDUS257	2.
@00262	DEDUS202	1.	@00360	DEDUS258	1.
@00263	DEDUS203	1.	@00361	DEDUS259	2.
@00264	DEDUS204	1.	@00363	DEDUS260	2.
@00265	DEDUT205	1.	@00365	DEDUS261	2.
@00266	DEDUT206	1.	@00367	DEDUS262	2.
@00267	DEDUT207	1.	@00369	DEDUS263	2.
@00268	DEDUT208	1.	@00371	DEDUS264	2.
@00269	DEDUT209	1.	@00373	DEDUS265	1.
@00270	DEDUT210	1.	@00374	DEDUS266	2.
@00271	DEDUT211	1.	@00376	DEDUS267	2.
@00272	DEDUM212	1.	@00378	DEDUS268	2.
@00273	DEDUM213	1.	@00380	DEDUS269	2.
@00274	DEDUM214	1.	@00382	DEDUS270	2.
@00275	DEDUM215	1.	@00384	DEDUS271	2.
@00276	DEDUM216	1.	@00386	DEDUT272	2.
@00277	DEDUM217	1.	@00388	DEDUS273	2.
@00278	DEDUT218	2.	@00390	DEDUT274	2.
@00280	DEDUT219	2.	@00392	DEDUT275	2.
@00282	DEDUT220	2.	@00394	DEDUT276	2.
@00284	DEDUT221	2.	@00396	DEDUT277	2.
@00286	DEDUS222	2.	@00398	DEDUT278	2.
@00288	DEDUS223	2.	@00400	DEDUT279	2.
@00290	DEDUT224	2.	@00402	DEDUT280	2.
@00292	DEDUT225	2.	@00404	DEDUT281	2.
@00294	DEDUT226	2.	@00406	DEDUT282	2.
@00296	DEDUS227	3.	@00408	DEDUT283	2.
@00299	DEDUS228	3.	@00410	DEDUT284	2.
@00302	DEDUS229	3.	@00412	DEDUT285	2.
@00305	DEDUS230	2.	@00414	DEDUT286	2.
@00307	DEDUS231	2.	@00416	DEDUT287	2.
@00309	DEDUT232	2.	@00418	DEDUT288	2.
@00311	DEDUT233	2.	@00420	DEDUT289	2.
@00313	DEDUT234	2.	@00422	DEDUT290	2.
@00315	DEDUT235	2.	@00424	DEDUT291	2.
@00317	DEDUT236	2.	@00426	DEDUT292	2.
@00319	DEDUT237	2.	@00428	DEDUT293	2.
@00321	DEDUT238	2.	@00430	DEDUT294	2.
@00323	DEDUT239	2.	@00432	DEDUT295	2.
@00325	DEDUT240	2.	@00434	DEDUT296	2.
@00327	DEDUT241	2.	@00436	DEDUT297	1.
@00329	DEDUT242	2.	@00437	DEDUT298	2.
@00331	DEDUT243	2.	@00439	DEDUT299	3.
@00333	DEDUT244	2.	@00442	DEDUT300	3.
@00335	DEDUM245	2.	@00445	DEDUT301	3.
@00337	DEDUM246	2.	@00448	DEDUT302	2.
@00339	DEDUM247	2.	@00450	DEDUT303	2.

@00452	DEDUT304	2.	@00510	DEDUP363	1.
@00454	DEDUM305	2.	@00511	DEDUP364	1.
@00456	DEDUS306	1.	@00512	DEDUP365	1.
@00457	DEDUS307	1.	@00513	DEDUP366	1.
@00458	DEDUS308	1.	@00514	DEDUP367	1.
@00459	DEDUS309	1.	@00515	DEDUP368	1.
@00460	DEDUS310	1.	@00516	DEDUP369	1.
@00461	DEDUS311	1.	@00517	DEDUP370	1.
@00462	DEDUS312	1.	@00518	DEDUP371	1.
@00463	DEDUS313	1.	@00519	DEDUP372	1.
@00464	DEDUS314	1.	@00520	DEDUP373	1.
@00465	DEDUS315	1.	@00521	DEDUP374	1.
@00466	DEDUS316	1.	@00522	DEDUP375	1.
@00467	DEDUS317	1.	@00523	DEDUP376	1.
@00468	DEDUT318	1.	@00524	DEDUP377	1.
@00469	DEDUT319	1.	@00525	DEDUP378	1.
@00470	DEDUT320	1.	@00526	DEDUP379	1.
@00471	DEDUT321	1.	@00527	DEDUP380	1.
@00472	DEDUT322	1.	@00528	DEDUP381	1.
@00473	DEDUT323	1.	@00529	DEDUP382	1.
@00474	DEDUT324	1.	@00530	DEDUP383	1.
@00475	DEDUT325	1.	@00531	DEDUP384	4.
@00476	DEDUT326	1.	@00535	DEDUP385	3.
@00477	DEDUT327	1.	@00538	DEDUP386	3.
@00478	DEDUT328	1.	@00541	DEDUP387	3.
@00479	DEDUT329	1.	@00544	DEDUP388	3.
@00480	DEDUT330	1.	@00547	DEDUP389	3.
@00481	DEDUT331	1.	@00550	DEDUP390	3.
@00482	DEDUT332	1.	@00553	DEDUP391	2.
@00483	DEDUT333	1.	@00555	DEDUP392	2.
@00484	DEDUT334	1.	@00557	DEDUP393	2.
@00485	DEDUT335	1.	@00559	DEDUP394	2.
@00486	DEDUT337	1.	@00561	DEDUP395	2.
@00487	DEDUT338	1.	@00563	DEDUP396	2.
@00488	DEDUT339	1.	@00565	DEDUP397	2.
@00489	DEDUT340	1.	@00567	DEDUP398	2.
@00490	DEDUT342	1.	@00569	DEDUP399	2.
@00491	DEDUP344	1.	@00571	DEDUP400	2.
@00492	DEDUP345	1.	@00573	DEDUP401	2.
@00493	DEDUP346	1.	@00575	DEDUP402	2.
@00494	DEDUP347	1.	@00577	DEDUP403	2.
@00495	DEDUP348	1.	@00579	DEDUP404	2.
@00496	DEDUP349	1.	@00581	DEDUP405	2.
@00497	DEDUP350	1.	@00583	DEDUP406	3.
@00498	DEDUP351	1.	@00586	DEDUP407	3.
@00499	DEDUP352	1.	@00589	DEDUP408	3.
@00500	DEDUP353	1.	@00592	DEDUP410	3.
@00501	DEDUP354	1.	@00595	DEDUP411	3.
@00502	DEDUP355	1.	@00598	DEDUP412	3.
@00503	DEDUP356	1.	@00601	DEDUP413	3.
@00504	DEDUP357	1.	@00604	DEDUP414	3.
@00505	DEDUP358	1.	@00607	DEDUP415	3.
@00506	DEDUP359	1.	@00610	DEDUP416	3.
@00507	DEDUP360	1.	@00613	DEDUP417	3.
@00508	DEDUP361	1.	@00616	DEDUP418	3.
@00509	DEDUP362	1.	@00619	DEDUP419	3.

@00622	DEDUP420	3.	@00717	DEDUP483	1.
@00625	DEDUP421	3.	@00718	DEDUP484	1.
@00628	DEDUP422	3.	@00719	DEDUP485	3.
@00631	DEDUP423	2.	@00722	DEDUP486	3.
@00633	DEDUP424	2.	@00725	DEDUP487	3.
@00635	DEDUP425	2.	@00728	DEDUP488	3.
@00637	DEDUP426	2.	@00731	DEDUP489	3.
@00639	DEDUP427	2.	@00734	DEDUP490	3.
@00641	DEDUP428	2.	@00737	DEDUP491	2.
@00643	DEDUP429	2.	@00739	DEDUP492	2.
@00645	DEDUP430	2.	@00741	DEDUP493	2.
@00647	DEDUP431	2.	@00743	DEDUP494	2.
@00649	DEDUP432	2.	@00745	DEDUP495	2.
@00651	DEDUP433	2.	@00747	DEDUP496	2.
@00653	DEDUP435	2.	@00749	DEDUP497	2.
@00655	DEDUP436	2.	@00751	DEDUP498	2.
@00657	DEDUP438	3.	@00753	DEDUP499	2.
@00660	DEDUP439	3.	@00755	DEDUP500	2.
@00663	DEDUP441	3.	@00757	DEDUP501	2.
@00666	DEDUP442	3.	@00759	DEDUP502	1.
@00669	DEDUP444	3.	@00760	DEDUP503	2.
@00672	DEDUP445	3.	@00762	DEDUP504	3.
@00675	DEDUP447	3.	@00765	DEDUP505	3.
@00678	DEDUP448	3.	@00768	DEDUP506	3.
@00681	DEDUP450	3.	@00771	DEDUP507	3.
@00684	DEDUP451	3.	@00774	DEDUP508	3.
@00687	DEDUP453	1.	@00777	DEDUP509	3.
@00688	DEDUP454	1.	@00780	DEDUP510	1.
@00689	DEDUP455	1.	@00781	DEDUP511	1.
@00690	DEDUP456	1.	@00782	DEDUP512	1.
@00691	DEDUP457	1.	@00783	DEDUP513	1.
@00692	DEDUP458	1.	@00784	DEDUP514	1.
@00693	DEDUP459	1.	@00785	DEDUP515	1.
@00694	DEDUP460	1.	@00786	DEDUP516	1.
@00695	DEDUP461	1.	@00787	DEDUP517	1.
@00696	DEDUP462	1.	@00788	DEDUP518	1.
@00697	DEDUP463	1.	@00789	DEDUP519	1.
@00698	DEDUP464	1.	@00790	DEDUP520	1.
@00699	DEDUP465	1.	@00791	DEPCS15	2.
@00700	DEDUP466	1.	@00793	DETSS17	2.
@00701	DEDUP467	1.	@00795	DETSS26	2.
@00702	DEDUP468	1.	@00797	DETSS28A	2.
@00703	DEDUP469	1.	@00799	DETSS28C	2.
@00704	DEDUP470	1.	@00801	DETSS50A	2.
@00705	DEDUP471	1.	@00803	DETSS50B	2.
@00706	DEDUP472	1.	@00805	DETSS51	2.
@00707	DEDUP473	1.	@00807	DGEHD03	2.
@00708	DEDUP474	1.	@00809	DMMPQ01	3.
@00709	DEDUP475	1.	@00812	DMMPQ02	\$1.
@00710	DEDUP476	1.	@00813	DMMPQ03A	4.
@00711	DEDUP477	1.	@00817	DMMPQ03B	2.
@00712	DEDUP478	1.	@00819	DMMPQ03C	2.
@00713	DEDUP479	1.	@00821	DMMPQ04	2.
@00714	DEDUP480	1.	@00823	DMMSQ01	3.
@00715	DEDUP481	1.	@00826	DMMSQ02	\$1.
@00716	DEDUP482	1.	@00827	DMMSQ03A	4.

@00831	DMMSQ03B	2.	@00953	DSPHS01	2.
@00833	DMMSQ03C	2.	@00955	DSDPQ1	2.
@00835	DMMSQ04	2.	@00957	DSDPQ3	4.
@00837	DMMCQ01	3.	@00961	DSDSQ1	2.
@00840	DMMCQ02	\$1.	@00963	DSDSQ3	4.
@00841	DMMCQ03A	4.	@00967	DSDPD01	3.
@00845	DMMCQ03B	2.	@00970	DSDPD02	3.
@00847	DMMCQ03C	2.	@00973	DSDPD03	2.
@00849	DDMCD03	2.	@00975	DSDPD04	2.
@00851	DDMCD06	2.	@00977	DSDPD05	2.
@00853	DDMCD06B	2.	@00979	DSDSD01	3.
@00855	DDMCD15	1.	@00982	DSDSD02	3.
@00856	DDMCD16	1.	@00985	DSDSD03	2.
@00857	DDMCD18	2.	@00987	DSDSD04	2.
@00859	DDMCD19	2.	@00989	DSDSD05	2.
@00861	DEDPdQ4b	2.	@00991	DEDCdQ0	2.
@00863	DEDSdQ4b	2.	@00993	DEDCcQ0A	2.
@00865	DEDPD01	2.	@00995	DEDCQ06	1.
@00867	DESD01	2.	@00996	DEDCD03	2.
@00869	DEDHcQ8A	1.	@00998	DEDCdQ7F	2.
@00870	DEDHcQ8B	1.	@01000	DEDCdQ9A	1.
@00871	DEDHcQ8C	1.	@01001	DEDCdQ9B	2.
@00872	DEDHcQ8D	1.	@01003	DEDCQ10	2.
@00873	DEDHcQ8E	1.	@01005	DEDCd11A	2.
@00874	DLFPcD5A	4.	@01007	DEDCQ14A	2.
@00878	DLFPcD6A	\$4.	@01009	DEDCb14A	2.
@00882	DLFPcD7A	2.	@01011	DEDCQ14B	2.
@00884	DLFPcD8A	2.	@01013	DEDCQ14C	2.
@00886	DLFPD25	1.	@01015	DEDCb14C	2.
@00887	DLFPD34	2.	@01017	DEDCQ14D	2.
@00889	DLFPbD38	2.	@01019	DEDCb14H	2.
@00891	DLFPD51	1.	@01021	DEDCd15C	1.
@00892	DLFScD5A	4.	@01022	DEDCd15D	2.
@00896	DLFScD6A	\$4.	@01024	DEDCQ17	2.
@00900	DLFScD7A	2.	@01026	DEDCQ18A	1.
@00902	DLFScD8A	2.	@01027	DEDCQ18B	2.
@00904	DLFSD25	1.	@01029	DEDCc18C	2.
@00905	DLFSD34	2.	@01031	DEDCc18D	2.
@00907	DLFSbD38	2.	@01033	DHLCcD2A	4. 2
@00909	DLFSD51	1.	@01037	DHLCbD4C	5. 3
@00910	DLFHD49B	2.	@01042	DHLCbD4D	1.
@00912	DLFHD50	2.	@01043	DHLCbD45	1.
@00914	DINHQ03	7.	@01044	DBECdS06	2.
@00921	DINPD02	2.	@01046	DBECdS07	2.
@00923	DINPCD04	7.	@01048	DBECdS08	2.
@00930	DINScD02	2.	@01050	DBECdS09	2.
@00932	DINScD04	7.	@01052	DBECS10	2.
@00939	DINH01B	2.	@01054	DBECdS11	2.
@00941	DCHPD01	1.	@01056	DSLcdQ7	2.
@00942	DRSPdD01	1.	@01058	DPRCS03	2.
@00943	DCHSD01	1.	@01060	DPRCS04	2.
@00944	DRSSdD01	1.	@01062	DPRCS05	2.
@00945	DDPPS01	2.	@01064	DPRCS06	2.
@00947	DFNHS01	2.	@01066	DPRCbS09	2.
@00949	DSFHS5	2.	@01068	DSDCQ1	2.
@00951	DSFHS6	2.	@01070	DSDCQ3	4.

@01074	DSDCD01	3.	@01124	DPRCQ07	2.
@01077	DSDCD02	3.	@01126	DPRCQ11	2.
@01080	DSDCD03	2.	@01128	CPPCS01	3.
@01082	DSDCD04	2.	@01131	CPPCS02	3.
@01084	DSDCD05	2.	@01134	CMACS01	2.
@01086	DPPCS01	3.	@01136	CMACS02	3.
@01089	DPPCS02	3.	@01139	CRECS01	2.
@01092	DMACS01	2.	@01141	CRECS02	3.
@01094	DMACS02	3.	@01144	BPPCS01	3.
@01097	DMACS03	3.	@01147	BPPCS02	3.
@01100	DGEHbD04	2.	@01150	BMACS01	2.
@01102	DDMCD08	2.	@01152	BMACS02	3.
@01104	DLFPQ02	1.	@01155	BRECS01	2.
@01105	DI NHdQ05	1.	@01157	BRECS02	3.
@01106	DHLPQ01	2.	@01160	APPCS01	3.
@01108	DHLPQ02	1.	@01163	APPCS02	3.
@01109	DSDPQ8	2.	@01166	AMACS01	2.
@01111	DHHHQ01	1.	@01168	AMACS02	3.
@01112	DHHHQ06	2.	@01171	FLAG_TE	1.
@01114	DHLCQ01	2.	@01172	FLAG_PR	1.
@01116	DHLCQ45I	1.	@01173	PERSRUK	\$14.
@01117	DHLCQ52A	1.	@01187	FIELDRUK	\$12.
@01118	DRLCQ01	2.	@01199	DSCH_ID	8.
@01120	DLTCdQ14	2.	@01207	CHILDIID	6.
@01122	DPRCQ03	2.	@01213	DEDUW01C	12. 4;

18.2 Secondary File

@00001	DAGCdQ1	1.	@00028	DAGCd202	1.
@00002	DAGCdQ2	1.	@00029	DAGCd204	1.
@00003	DAGCdQ3	1.	@00030	DAGCd205	1.
@00004	DAGCdQ4	1.	@00031	DAGCd206	1.
@00005	DAGCdQ5	1.	@00032	DAGCd207	1.
@00006	DAGCdQ6	1.	@00033	DAGCd261	1.
@00007	DAGCdQ33	1.	@00034	DAGCd262	1.
@00008	DAGCdQ34	1.	@00035	DAGCd263	1.
@00009	DAGCdQ35	1.	@00036	DAGCd310	1.
@00010	DAGCdQ36	1.	@00037	DAGCd311	1.
@00011	DAGCdQ37	1.	@00038	DAGCd312	1.
@00012	DAGCdQ38	1.	@00039	DAGCd352	1.
@00013	DAGCdQ39	1.	@00040	DAGCd353	1.
@00014	DAGCdQ77	1.	@00041	DAGCd387	1.
@00015	DAGCdQ79	1.	@00042	DAGCd388	1.
@00016	DAGCdQ80	1.	@00043	DAGCd389	1.
@00017	DAGCdQ81	1.	@00044	DAGCd390	1.
@00018	DAGCdQ82	1.	@00045	DAGCd393	1.
@00019	DAGCdQ83	1.	@00046	DAGCd394	1.
@00020	DAGCdQ84	1.	@00047	DAGCd430	1.
@00021	DAGCd140	1.	@00048	DAGCd432	1.
@00022	DAGCd141	1.	@00049	DAGCdQ7	1.
@00023	DAGCd142	1.	@00050	DAGCdQ8	1.
@00024	DAGCd143	1.	@00051	DAGCdQ9	1.
@00025	DAGCd144	1.	@00052	DAGCdQ10	1.
@00026	DAGCd145	1.	@00053	DAGCdQ11	1.
@00027	DAGCd146	1.	@00054	DAGCdQ12	1.

@00055	DAGCdQ40	1.	@00111	DAGCdQ50	1.
@00056	DAGCdQ41	1.	@00112	DAGCdQ51	1.
@00057	DAGCdQ42	1.	@00113	DAGCdQ52	1.
@00058	DAGCdQ43	1.	@00114	DAGCdQ53	1.
@00059	DAGCdQ44	1.	@00115	DAGCdQ54	1.
@00060	DAGCdQ45	1.	@00116	DAGCdQ55	1.
@00061	DAGCdQ46	1.	@00117	DAGCd101	1.
@00062	DAGCdQ47	1.	@00118	DAGCd102	1.
@00063	DAGCdQ89	1.	@00119	DAGCd103	1.
@00064	DAGCdQ91	1.	@00120	DAGCd104	1.
@00065	DAGCdQ92	1.	@00121	DAGCd105	1.
@00066	DAGCdQ93	1.	@00122	DAGCd106	1.
@00067	DAGCdQ94	1.	@00123	DAGCd107	1.
@00068	DAGCdQ95	1.	@00124	DAGCd108	1.
@00069	DAGCdQ96	1.	@00125	DAGCd166	1.
@00070	DAGCd150	1.	@00126	DAGCd167	1.
@00071	DAGCd151	1.	@00127	DAGCd168	1.
@00072	DAGCd152	1.	@00128	DAGCd169	1.
@00073	DAGCd153	1.	@00129	DAGCd170	1.
@00074	DAGCd154	1.	@00130	DAGCd171	1.
@00075	DAGCd155	1.	@00131	DAGCd224	1.
@00076	DAGCd156	1.	@00132	DAGCd226	1.
@00077	DAGCd157	1.	@00133	DAGCd227	1.
@00078	DAGCd158	1.	@00134	DAGCd228	1.
@00079	DAGCd159	1.	@00135	DAGCd229	1.
@00080	DAGCd160	1.	@00136	DAGCd230	1.
@00081	DAGCd212	1.	@00137	DAGCd231	1.
@00082	DAGCd213	1.	@00138	DAGCd232	1.
@00083	DAGCd215	1.	@00139	DAGCd277	1.
@00084	DAGCd216	1.	@00140	DAGCd282	1.
@00085	DAGCd217	1.	@00141	DAGCd283	1.
@00086	DAGCd218	1.	@00142	DAGCd284	1.
@00087	DAGCd272	1.	@00143	DAGCd302	1.
@00088	DAGCd273	1.	@00144	DAGCd327	1.
@00089	DAGCd274	1.	@00145	DAGCd328	1.
@00090	DAGCd318	1.	@00146	DAGCd368	1.
@00091	DAGCd320	1.	@00147	DAGCd369	1.
@00092	DAGCd321	1.	@00148	DAGCd403	1.
@00093	DAGCd359	1.	@00149	DAGCd404	1.
@00094	DAGCd360	1.	@00150	DAGCd405	1.
@00095	DAGCd361	1.	@00151	DAGCd406	1.
@00096	DAGCd397	1.	@00152	DAGCd407	1.
@00097	DAGCd398	1.	@00153	DAGCd408	1.
@00098	DAGCd399	1.	@00154	DAGCd409	1.
@00099	DAGCd400	1.	@00155	DAGCd410	1.
@00100	DAGCd401	1.	@00156	DAGCd442	1.
@00101	DAGCd437	1.	@00157	DAGCd443	1.
@00102	DAGCd438	1.	@00158	DAGCd444	1.
@00103	DAGCdQ13	1.	@00159	DAGCdQ19	1.
@00104	DAGCdQ14	1.	@00160	DAGCdQ20	1.
@00105	DAGCdQ15	1.	@00161	DAGCdQ21	1.
@00106	DAGCdQ16	1.	@00162	DAGCdQ22	1.
@00107	DAGCdQ17	1.	@00163	DAGCdQ23	1.
@00108	DAGCdQ18	1.	@00164	DAGCdQ24	1.
@00109	DAGCdQ48	1.	@00165	DAGCdQ56	1.
@00110	DAGCdQ49	1.	@00166	DAGCdQ57	1.

@00167	DAGCdQ58	1.	@00220	DAGCdQ27	1.
@00168	DAGCdQ59	1.	@00221	DAGCdQ28	1.
@00169	DAGCdQ60	1.	@00222	DAGCdQ29	1.
@00170	DAGCdQ61	1.	@00223	DAGCdQ30	1.
@00171	DAGCdQ62	1.	@00224	DAGCdQ65	1.
@00172	DAGCdQ63	1.	@00225	DAGCdQ66	1.
@00173	DAGCd113	1.	@00226	DAGCdQ67	1.
@00174	DAGCd114	1.	@00227	DAGCdQ68	1.
@00175	DAGCd115	1.	@00228	DAGCdQ69	1.
@00176	DAGCd116	1.	@00229	DAGCdQ70	1.
@00177	DAGCd117	1.	@00230	DAGCdQ71	1.
@00178	DAGCd118	1.	@00231	DAGCd126	1.
@00179	DAGCd119	1.	@00232	DAGCd127	1.
@00180	DAGCd120	1.	@00233	DAGCd128	1.
@00181	DAGCd174	1.	@00234	DAGCd129	1.
@00182	DAGCd175	1.	@00235	DAGCd130	1.
@00183	DAGCd176	1.	@00236	DAGCd131	1.
@00184	DAGCd177	1.	@00237	DAGCd132	1.
@00185	DAGCd178	1.	@00238	DAGCd187	1.
@00186	DAGCd179	1.	@00239	DAGCd188	1.
@00187	DAGCd180	1.	@00240	DAGCd190	1.
@00188	DAGCd181	1.	@00241	DAGCd191	1.
@00189	DAGCd182	1.	@00242	DAGCd192	1.
@00190	DAGCd183	1.	@00243	DAGCd193	1.
@00191	DAGCd235	1.	@00244	DAGCd194	1.
@00192	DAGCd236	1.	@00245	DAGCd195	1.
@00193	DAGCd237	1.	@00246	DAGCd196	1.
@00194	DAGCd238	1.	@00247	DAGCd197	1.
@00195	DAGCd239	1.	@00248	DAGCd249	1.
@00196	DAGCd240	1.	@00249	DAGCd250	1.
@00197	DAGCd241	1.	@00250	DAGCd251	1.
@00198	DAGCd285	1.	@00251	DAGCd252	1.
@00199	DAGCd287	1.	@00252	DAGCd298	1.
@00200	DAGCd288	1.	@00253	DAGCd299	1.
@00201	DAGCd289	1.	@00254	DAGCd300	1.
@00202	DAGCd290	1.	@00255	DAGCd301	1.
@00203	DAGCd291	1.	@00256	DAGCd341	1.
@00204	DAGCd292	1.	@00257	DAGCd342	1.
@00205	DAGCd333	1.	@00258	DAGCd343	1.
@00206	DAGCd334	1.	@00259	DAGCd344	1.
@00207	DAGCd336	1.	@00260	DAGCd384	1.
@00208	DAGCd337	1.	@00261	DAGCd385	1.
@00209	DAGCd376	1.	@00262	DAGCd420	1.
@00210	DAGCd377	1.	@00263	DAGCd422	1.
@00211	DAGCd413	1.	@00264	DAGCd423	1.
@00212	DAGCd414	1.	@00265	DAGCd424	1.
@00213	DAGCd416	1.	@00266	DAGCd425	1.
@00214	DAGCd417	1.	@00267	FI ELDRUK	\$12.
@00215	DAGCd418	1.	@00279	PERSRUK	\$14.
@00216	DAGCd448	1.	@00293	CHI LDID	6.
@00217	DAGCd450	1.	@00299	DWTCW01C	12. 4
@00218	DAGCdQ25	1.	@00311	DWTCW01L	12. 4
@00219	DAGCdQ26	1.	@00323	DWTCWd1L	12. 4;