

The National Longitudinal Survey of Children and Youth – Overview and Changes After Three Cycles

Sylvie Michaud
Special Surveys Division
Statistics Canada
Ottawa, Ontario, Canada

Abstract:

The National Longitudinal Survey of Children and Youth is a long-term study to monitor child development and well being of Canada's children as they grow from infancy to adulthood. To do so, a representative sample of Canadian children aged between 0-11 years old was selected and interviewed in 1994-1995. Interviews are conducted every two years and the current plans are to follow that cohort of children until they reach the age of 25. The survey has now gone through three collection cycles and already a number of changes have been observed. The paper will give an overview of the objectives of the survey, the survey design, the collection methodology, the survey content and the products and research that has already been done on the survey. The last section will present the future direction of the survey.

Résumé:

L'Enquête nationale longitudinale sur les enfants et les jeunes concerne le développement et le bien-être des jeunes de l'enfance à la vie adulte. Un échantillon représentatif d'enfants des âges 0-11 ans a d'abord fait parti de l'enquête en 1994-1995. On veut continuer avec un interview chaque deux ans jusqu'à ce que les enfants ont 25 ans. Sur les trois premiers cycles, il y a déjà plusieurs changements. On décrit ici les objectifs de l'enquête, son dessin, la méthodologie de collecte, le contenu, la recherche qui à déjà été faite, et le future de l'enquête.

Key words: Child development, Human Resources Development Canada, NLSCY survey design, content themes, public-use microdata files

Introduction

The National Longitudinal Survey of Children and Youth (NLSCY) is a long-term study conducted in partnership by Human Resources Development Canada (HRDC) and Statistics Canada. The primary objective of the NLSCY is to monitor the development and well being of Canada's children as they grow from infancy to adulthood. More specifically, the objectives of the NLSCY are:

- to determine the prevalence of various biological, social and economic characteristics and risk factors among children and youth;
- to support our understanding of the determinants of child development and well being and of the pathways of their influence on child outcomes; and
- to provide this information to policy and program officials for use in developing effective policies and strategies to help young people live healthy, active and rewarding lives.

Statistics Canada is responsible for data collection, while HRDC provides overall direction to the survey. Both agencies play a role in funding, development of survey content, research and dissemination of findings. In addition, HRDC and Statistics Canada continue to benefit from the advice and input of a variety of other interested partners. Other federal departments are kept informed of developments and provide advice on the survey via the NLSCY Interdepartmental Committee. Provincial and territorial governments have been kept informed of progress and have provided input via representatives in each province and territory.

*The National Longitudinal Survey of Children and Youth –
Overview and Changes after Three Cycles*

HRDC and Statistics Canada have relied heavily on advice provided by HRDC's Expert Advisory Group on Children and Families (EAG), a multidisciplinary group of Canadian and international experts in child development. These experts are consulted on survey design, survey questions and research priorities, and who are responsible for much of the original research that has been conducted using the survey data. Additional experts are consulted on an ad hoc basis as required.

Survey Design

Initial Sample Design

The goal of the NLSCY is to follow, until adulthood, a representative sample of Canadian children who were aged from new-borns to 11 years during the first cycle in 1994-95, and to collect information about these children every two years. The NLSCY target population for the first data collection in 1994-95 consisted of the population of Canadian children new-borns to 11 years of age who lived in private households. Children living in institutions and children living on Indian reserves were not targeted by the survey, as the households for the main component were selected from Statistics Canada's Labour Force Survey (LFS) sample frame. In the first cycle, the children in each economic family¹ were selected at random, up to a maximum of four children per household. [In cycle 2, only two children were followed up because of response burden on the households with a large number of children]

By definition, the first cycle of the survey provided cross-sectional or prevalence information only. However, the NLSCY will be able to provide cross-sectional and longitudinal information. The longitudinal data are central to tracking developmental changes in children over time and to studying the influence of children's environments. The cross-sectional and longitudinal estimates will be produced at the national level for each age group. They will be produced at the provincial level for children from aggregated age groups to the extent possible, depending on the size of the sample and the reliability of the estimates.

The National Population Health Survey is another national longitudinal survey being conducted by Statistics Canada in the same years as the NLSCY. Its purpose is to produce reliable estimates of the current physical and mental health of Canadian residents of all ages and to identify the factors that determine good and ill health. Because both the NLSCY and the NPHS needed to collect data on the health of Canadian children, it was initially decided that a portion of the sample and content of the two surveys would be integrated in the provinces. The children selected by the NPHS were part of the sample for both surveys in cycle 1. For the following collection cycles however, to limit response burden,

these children will be surveyed through the NPHS only (this represents 3,844 children for longitudinal survey).

Maintaining Cross-sectional Representation

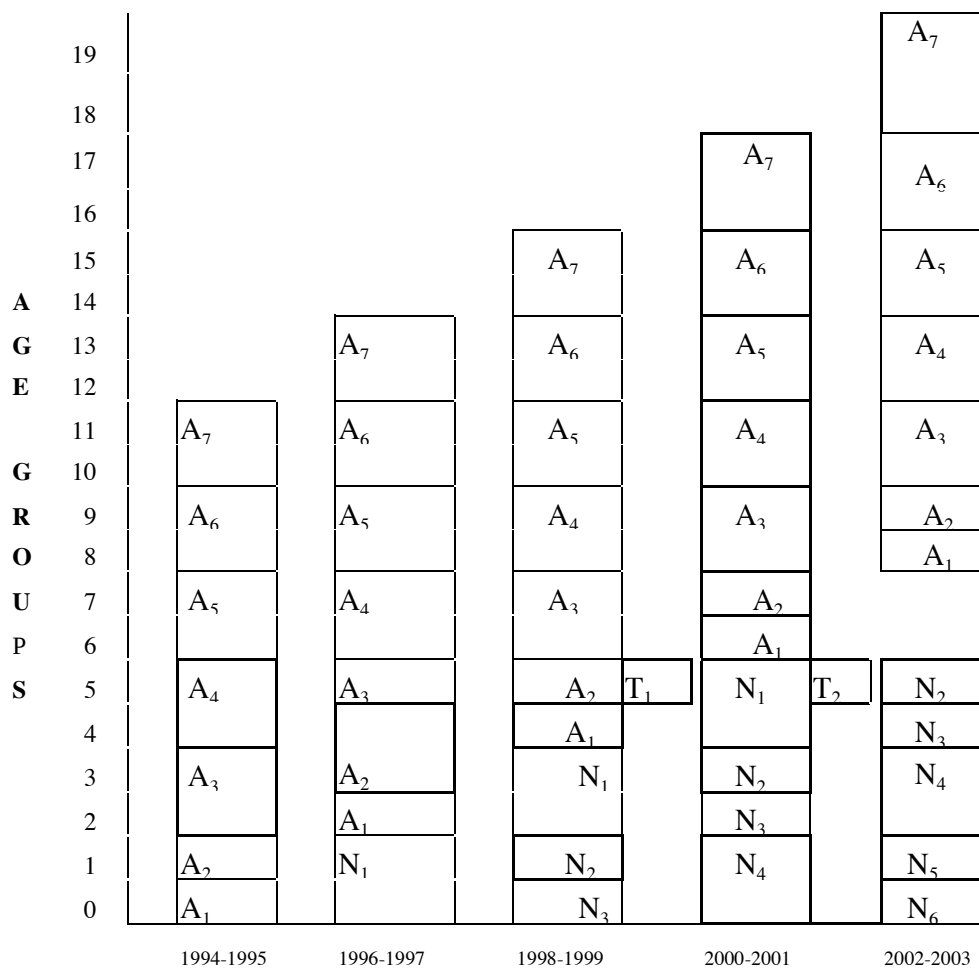
In the second and third collection cycles, the NLSCY target population for the purposes of longitudinal estimates retained the children from the respondent households for the first cycle, aged 2 to 13 years in cycle 2 and 4 to 15 in cycle 3. For the purposes of producing cross-sectional estimates, a sample of children new-borns to 1 year was selected for each new cycle (the current plans are to interview them every two years until they reach the age of 5) since these ages were no longer covered by the longitudinal sample. In cycle 2, these children were from households selected from the LFS sample frame, as well as from NLSCY longitudinal households with children born between the first and second data collections.

In 1997, the Speech from the Throne mentioned the commitment from the federal government to measure and report regularly on the readiness to learn of Canada's children so that Canada can assess its progress in providing children with the best possible start. So in cycle 3, in addition to the sample of new-borns to 1 year old for cross-sectional purposes, a sample of children 5 years old has been surveyed to allow estimation of provincial outcomes for 5 years old children. The 1-year old cohort was also increased in size. This should allow for provincial reporting of key outcomes every four years. Because of its size, the new cohorts of 1 and 5 years old were selected from administrative files in vital statistics. The cohort of new-borns was selected from the LFS sample frame. In cycle 3, siblings or new-borns since the last interview were not added to the sample like it was done in cycle 2.

Figure 1 provides an overview of the longitudinal and cross-sectional aspects of the NLSCY. A1 to A7 represent the longitudinal sample. N1 to N6 represent the additional cross-sectional sample for cycles 2 to 5. T1 represents an extra top-up sample of children of 5 years old that has been interviewed in cycle 3. Cross-sectional estimates will be produced for all age groups until cycle 4, e.g., for A1 to A7 for 1994-95, for N1 to A7 for 1996-97, for N3 to A7 for cycle 3 and for N4 to A7 in cycle 4. Starting in cycle 5, cross-sectional estimates should be possible only for children aged between 0 and 5. Longitudinal estimates will be produced starting with the second cycle to show changes in sample characteristics between collection cycles, e.g., between A3 in 1994-95 (children 2-3 years of age) and A3 in 1996-97 (the same children, now 4-5 years of age). Some longitudinal analysis will be possible on the younger age cohorts, but for a much shorter time period.

*The National Longitudinal Survey of Children and Youth –
Overview and Changes after Three Cycles*

Figure 1: NLSCY Sample Design (for the first five cycles)



COLLECTION CYCLES

This figure shows the first four data collection cycles for the NLSCY. The longitudinal sample are represented by the 'As' and the expanded cross-sectional sample by the 'Ns.' Boxes A1 to A7 represent the age groups used for the longitudinal sample: 2 years, 3 years, 4 to 5 years, 6 to 7 years, 8 to 9 years, 10 to 11 years and 12 to 13 years for the second collection cycle (1996-97). The supplementary sample from New Brunswick is not represented here.

In the cycles subsequent to cycle 1, it is intended that the NLSCY sample will be upgraded in age groups between 0 and 5 no longer covered by the longitudinal sample, to maintain coverage of the lower age ranges for cross-sectional purposes. The "N's" in Figure 1 represent the augmented portion of the cross-sectional sample. At the present time, it is planned that children added to upgrade the sample be followed longitudinally only until they reach the age of 5. This means that the top-up sample of 5 years old (labelled T) will be interviewed only once. For cycle 3, the additional sample was approximately 2,000 children 0 to 11 months. For children 1 year old, the sample was approximately 7,300. And approximately 6,400 extra children 5 years old have been added to the longitudinal sample.

Additional Provincial Buy-in

Provinces can request to buy an additional sample. In cycle 2, an additional sample of 500 children aged between 2 and 5 years old was added to the NLSCY.

Sample Allocation

The sample allocation for the first cycle of the NLSCY was based on several requirements. Sufficient sample was required in each age group to produce reliable estimates at the national level. There had to be a sufficient sample at the national level to reliably measure characteristics with a national prevalence of 4% for each age group after five cycles (min-p of 4%)². Secondly, a sufficient sample in each province and territory was required to produce reliable estimates for all children from new-borns to 11 years, with no age breakdown.

It should be noted that the NLSCY uses a cluster sample³, not a simple random sample. Thus, certain characteristics may not be measurable below the national level. The sample was first allocated by age group. A second step involved an allocation by province to make sure that the smaller provinces would have sufficient sample. Children of zero to 11 months and one-year of age were over-sampled by keeping them as separate groups.

Table 1 presents the longitudinal sample obtained for cycle 1, cycle 2 and cycle 3. In cycle 1, a response rate of 86% has been obtained. In cycle 2, only respondents to cycle 1 have been followed and this became the basis for the calculation of response rates in cycle 2. However, there has been a substantial decrease of the sample in cycle 2, in large part due to the fact that the sample that participated both in the NLSCY and NPHS was given back to NPHS. Therefore, the number of responding children is quite smaller. The response rate to cycle 3 is preliminary. Response codes are usually validated with actual

*The National Longitudinal Survey of Children and Youth –
Overview and Changes after Three Cycles*

response and the numbers could be somewhat different. Table 2 gives an overview of the information that has been collected from the various samples, by province.

Table 1
**Number of Longitudinal Children in the Longitudinal
Cohort by Age Group**

Age	Cycle 1 (# responding)	Cycle 2 (# responding after sub-sampling)	Cycle 3 (# responding)
0	2,227	–	–
1	2,469	–	–
2-3	3,909	3,654	–
4-5	3,728	2,697	3,506
6-7	3,550	2,429	2,587
8-9	3,514	2,169	2,317
10-11	3,434	2,249	2,086
12-13	–	2,270	2,129
14-15	–	–	2,174
Total	22,831*	15,468	14,799
Response Rate	86%	92% of Cycle 1 respondents	88% of Cycle 1 respondents

* 16,903 were kept after sub-sampling

Data Collection

The main NLSCY household data collection in the provinces was carried out using computer-assisted interviewing (CAI), either in person or by telephone. Statistics Canada interviewers conducted the interviews. Contacting the household and asking for demographic information was generally done by telephone. Interviews continued in person except for households with children aged between 0 and 3 years old only, for which interviews took place entirely by

Table 2
**Number of Responding Children in Cross-sectional and Supplementary
Samples in Cycle 3 Classified by Provinces**

Province	Long. Cohort	0-1 years old	2-3 years old	Extra 5 years old	Total
Newfoundland	804	587	91	199	1,681
P.E.I.	432	284	71	186	973
Nova Scotia	1,042	608	165	214	2,029
New Brunswick	937	603	226	221	1,987
Quebec	2,868	1,423	781	1,444	6,516
Ontario	3,745	2,051	1,139	1,836	8,771
Manitoba	1,075	661	303	213	2,252
Saskatchewan	1,210	630	269	193	2,302
Alberta	1,390	767	312	510	2,979
British Columbia	1,278	698	298	587	2,861
Total	14,781*	8,312	3,655	5,603	32,351

- these are preliminary numbers and data by province have not been cleaned or can be missing.

telephone in the majority of cases. This was feasible since the children were young and weren't required to complete a questionnaire or a test.

Face-to face interviewing for longitudinal households was used for several reasons. Personal interviewing helped secure the household's participation over a long period of time and allowed for the development of a rapport between the respondent and the interviewer. Also, the nature of some elements of the NLSCY made it necessary for collection of a portion of the data to occur through personal interviewing. For example, an interviewer-administered assessment to measure receptive vocabulary was conducted in the home for children aged 4 to 6 years of age. In addition, 10 to 15 year olds were asked to fill out a self-completed questionnaire, in most cases while the interview with the parent was being conducted. Children in second grade and up were given a reading and mathematical aptitude indicator. Furthermore, the parent's signed consent was required in order to collect information from teachers and principals and to allow children in grades 2 and above to be given a reading comprehension and mathematics computation exercise in school.

The National Longitudinal Survey of Children and Youth – Overview and Changes after Three Cycles

Once the household collection was completed, questionnaires were mailed to, and returned by, teachers and principals of school-aged children for whom parental consent had been given. Figure 2 presents the schema of the survey instruments.

The main respondent for cycle 3, as for cycle 1, was the person in the household most knowledgeable about the child(ren). In most cases, the respondent was the mother. This person provided information on herself, her spouse/partner, the selected child(ren) and household members. The child was the respondent for the direct assessment of receptive vocabulary (4, 5 and 6-years old and in first grade or lower); self-administered questionnaires for 10-11 years old, 12-13 years old and 14-15 years old; and the reading and mathematical aptitude indicator (children in second grade and over). These were administered in the home. They were also the respondent for the Reading Comprehension and Mathematics Computation Exercise (second grade and over) given to them in school. The interviewer completed an evaluation of the Peabody Picture Vocabulary Test (PPVT) conditions and of the indicator. The child's teacher and principal responded to the Teacher's and Principal's Questionnaires.

Data collection in the households takes place normally between November and March. This is done so that a school mail out can be done to teachers and principals. In cycle 3, the addition of extra sample for both 1 and 5 years old required extending the collection period, and this will likely be carried through in future cycles. The first collection period in households started in November. It collected mainly the information on the longitudinal sample (for the 6-15 years old mainly). The second collection started in February 1999 and the sample was composed mainly of children 4 and 5 years old (non-respondents from the November collection were also attempted again). The last collection started in April 1999 and collected mostly information from children 0-3 years old. Each collection lasted approximately two months. The school collection was carried out in the May and June of 1999.

Content

The NLSCY is a comprehensive survey that examines a variety of factors thought to influence child growth and development. Information is collected on the child's parent(s) and other family members, on the characteristics of the child's family and on the child's school. In addition, the NLSCY collects information on the child's health, development, temperament, behaviour, relationships, childcare and school experiences, participation in activities and family and custody history. The various survey instruments are presented in the documents **National Longitudinal Survey of Children and Youth: Survey Instruments Catalogue No. 89FOO77XPE**. Figure 3 provides a summary of the content of the third collection cycle of the NLSCY.

Figure 2. NLSCY Survey Instruments - Provinces

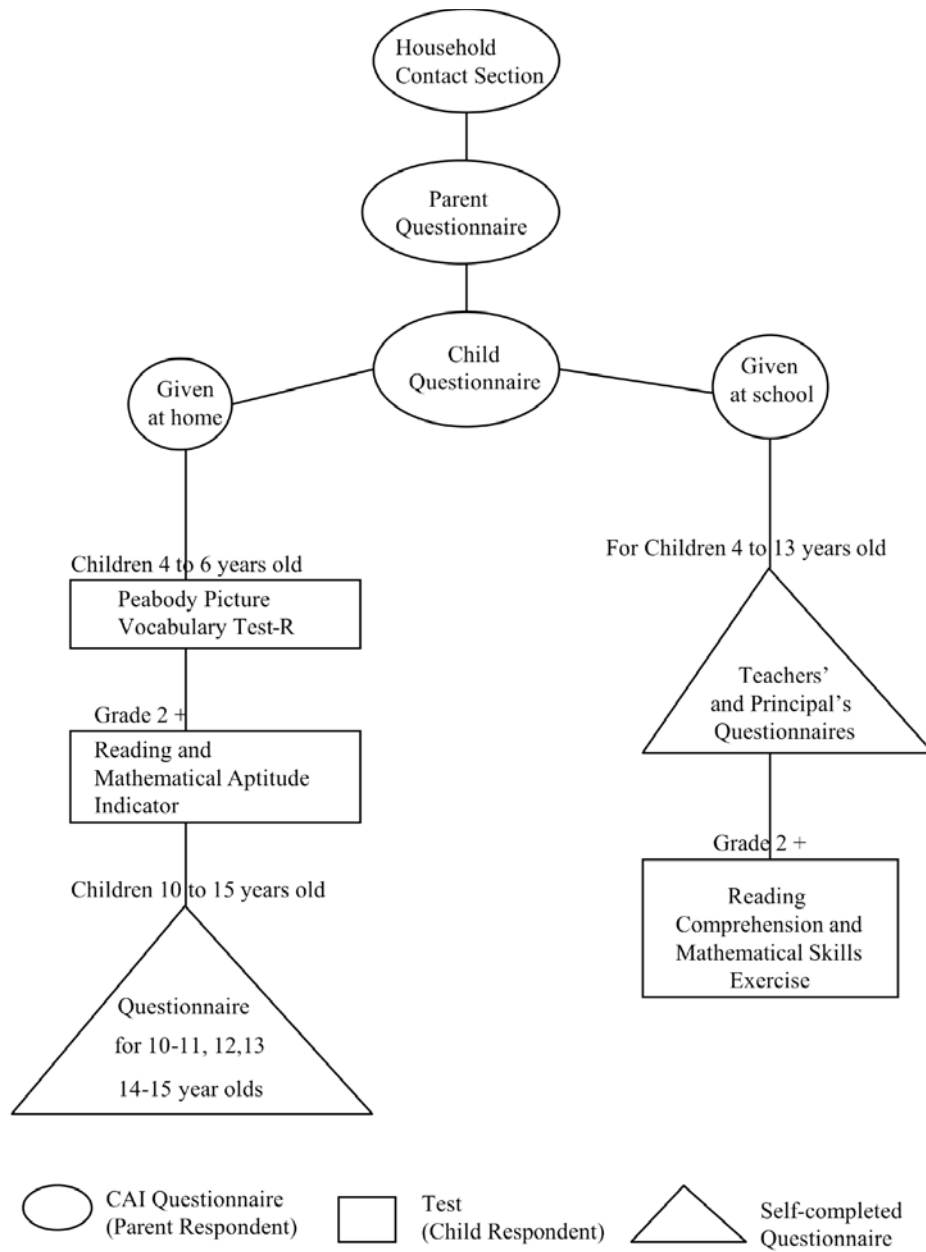
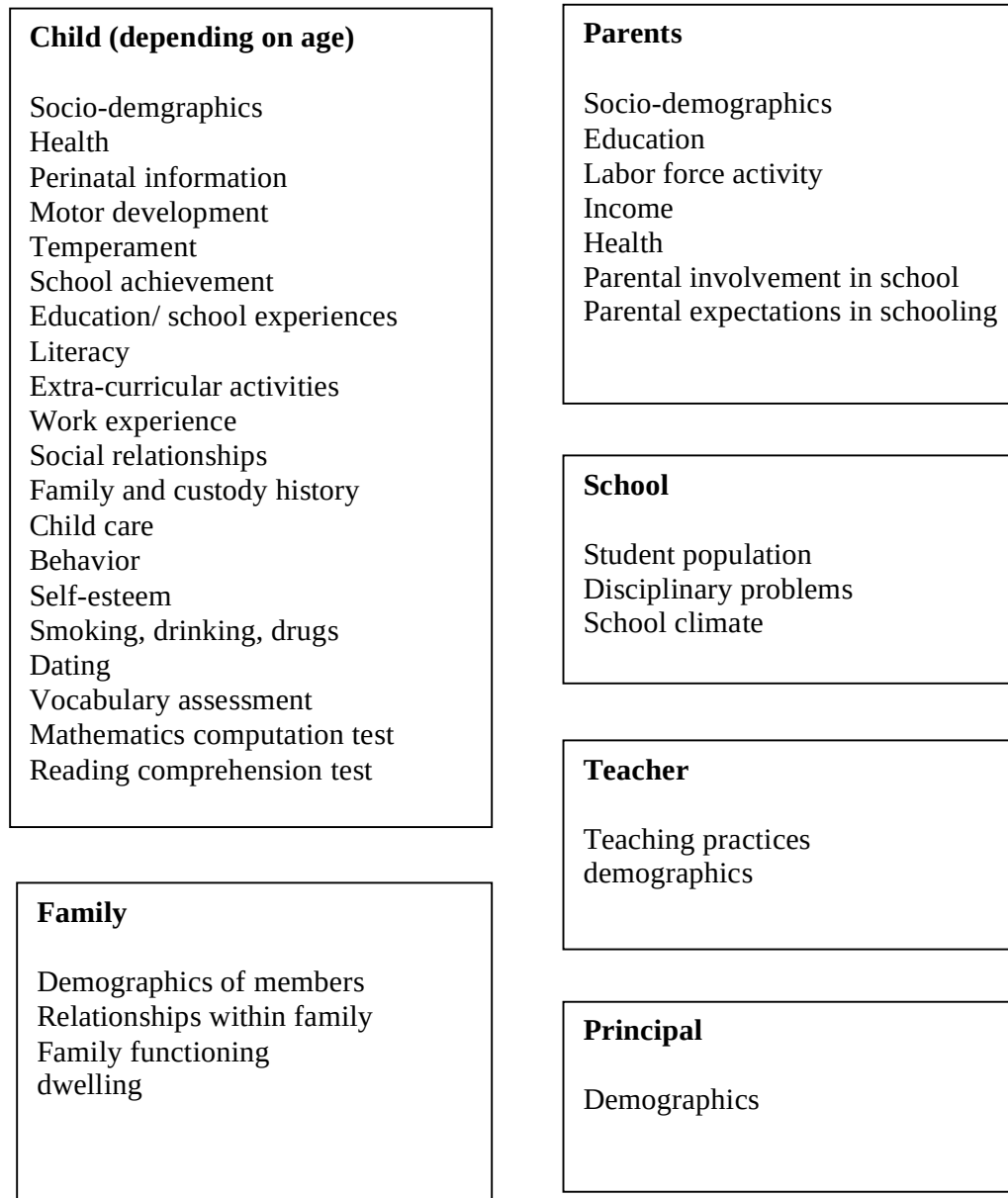


Figure 3: Content Overview



In general, an extra effort has been taken to maintain consistency in the content and the specific question wording throughout the first three cycles of the survey. Some content revisions are planned for cycle 4 and this will be highlighted in the section Next Steps below. There have been modifications through the cycles however.

In the home interview, in cycle 2, the PPVT was extended to children aged between 4 and 6. A small test in the home called the 'locator' was also administered to children in grade 2 or higher; this test was designed to try to determine the level of the test that should be administered to children in school. Since the test in school is short, and since curriculum is a provincial jurisdiction, it was difficult to design a test that could accommodate requirements for all provinces. [This had led to a ceiling effect in the tests in cycle 1.] The labour force section of the parent questionnaire was also modified and shortened. Some sections were dropped (neighbourhood and social support) while some sections were asked of only certain respondents. The dropped sections were re-introduced in cycle 3.

In cycle 2, a reading comprehension was added to the mathematics computation test in school. The teachers' questionnaires were also designed differently for kindergarten teachers, for teachers in a classroom environment with only one teacher versus multiple teachers.

The variable names have been constructed in a way that will hopefully make it easier for researchers to know when a variable or a new concept has been introduced for a given cycle. Collection names can also be different from output variable names. A concordance table facilitates identifying the collection names.

Statistics Canada Products

The data collected from the various cycles of the survey are being released by Statistics Canada in several "waves" due to the size of the information collected.

For the first collection cycle, the first wave of data was released in October 1996, along with a publication entitled "Growing Up in Canada", which presented the first research studies undertaken using NLSCY data. Further information from teachers and principals in cycle 1 was released in April 1997. The third wave of data was released in the spring of 1998 and it dealt with variables on family history and custody. A micro-data file for the north was also created.

The first release of the longitudinal data was in October 1998 at a conference called "Investing in Children" organised by Human Resources Development

*The National Longitudinal Survey of Children and Youth –
Overview and Changes after Three Cycles*

Canada. At that conference, a number of other papers were presented with further analysis done with the cycle 1 data. The second release of cycle 2 data was in July 1999, with the information provided by teachers and principals as well as information from the self-completed interviews. The Daily article in July focused on results from 12-13, while a release on October 14 gave some results on early childhood. The third release is planned for the fall 1999. Processing has started with the third collection cycle and a first release is planned for July 2000.

Micro-data Files

Three kinds of micro-data file are produced with each collection cycle; a master file, a shared file and a public use micro-data file. On all the files, the child is the unit of analysis and all information for a given child is combined into one single record.

The master file combines all the information related to a specific child. Because of the volume of information, this file actually contains three files (a primary, a secondary, and a tertiary file). At the moment, there is no longitudinal file as such. Researchers have to link by a child identifier to get longitudinal information. What is included in the primary, secondary and tertiary files have changed from cycle 1 and cycle 2. Starting in cycle 2, the information in the primary file should be similar to the content of the public use micro-data file.

At the end of the survey, respondents are asked for their permission to share the information with Human Resources Development Canada. Approximately 95% of respondents agree to share their information. For those cases, names and addresses are removed but the rest of the information is released in a file to HRDC called the shared file. HRDC is committed to protect the confidentiality of that information.

The third file is the public use micro-data file. The general public can buy this file, and it is also available through the data liberation initiative. In order to ensure respondent confidentiality, a longitudinal public use micro-data file is not available to the public. Only cross-sectional files are available and they can be purchased at each cycle. Certain variables are not available on the public use micro-data file. For example, the self-completed questionnaires or the responses from the teachers or the principals are not released on the public use micro-data file. This was done to limit the risk of disclosure that would happen where someone (the teacher, the teen, the parent), knowing what they reported in their questionnaire, could have access to someone else's response. Those wishing access to suppressed or longitudinal data can do so by either remote access or through custom tabulations. In cycle 2, a micro-data file has been released separately with the information from the self-reports. This was done because it was felt there could be interest in the answers from teens. Confidentiality checks

have been done to ensure that there could not be a link done between the two public use micro-data files.

Remote Access

Through remote access, researchers can have access to suppressed data by submitting programs to run on the NLSCY data set at Statistics Canada. A 'dummy' research file will be made available to researchers to check the logic and syntax of their programs. Researchers will transmit their programs electronically to Statistics Canada via the INTERNET, which will then be moved into the Department's internal, secure environment. Next, the code is processed on a PC, the results vetted for confidentiality, and shipped back to the client. It should be noted that the onus is with the user to submit retrieval programs that are correct and tested. Statistics Canada will review results only for confidentiality concerns and will not make any assessment whatsoever as to whether or not the submitted program has worked properly. Remote data access is available on a cost recovery basis. For more information please contact nlscy@statcan.ca.

Custom Tabulations

Statistics Canada retains a master micro-data file from which specific micro-data files and special tabulations can be requested by contacting nlscy@statcan.ca.

Next Steps

There are a few more revisions or modifications that can be expected for the NLSCY in cycle 4. The oldest cohort in NLSCY is turning 16-17 in cycle 4 and this will lead to changes. Among other things, starting in cycle 4, the 16-17 year old 'child' will become the person most knowledgeable. This means that most of the questions will be asked of him/her and not of the parents. A few questions will still be asked of the parents, related to labour market activity, income, education and parental expectations. Also, since at 16 or 17, young adults can finish or quit school, there will not be a contact with the teachers or the school principals. The test that used to be administered at school on mathematics computation and reading comprehension will be replaced by a small literacy test at home. In addition, there will likely be more direct measures that will be added to the sample of the young cohort.

Finally, a community project has been tested in North York in June 1999. The community project was done through a joint collaboration of HRDC and the Early Action Group that had a community initiative in North York. The purpose

*The National Longitudinal Survey of Children and Youth –
Overview and Changes after Three Cycles*

of this project is to try to measure the influence of the community on selected outcomes, and the project is to take place over a five years span. Every year, teachers are asked to fill in a questionnaire for all children in kindergarten from the protestant school board of North York. The questionnaire was developed by Dan Offord and Magdalena Janus and it gives some basic outcomes for those children. In years 1 and 5, a sample of those kindergarten children has also been selected and the parents are contacted and the NLSCY questionnaire is administered. This will allow linking some outcomes with a series of potential determinants. At the same time, more detailed information will be asked from the parents regarding resources utilisation in the communities. An ethnographic study will be done in years 1 and 5 to be able to determine resources availability. It is expected that the community project will be repeated in 5 more communities in early 2000.

As the NLSCY grows, it is hoped that it will be used more and more as a monitoring tool for child outcomes but also as a research vehicle to understand the links between various factors that may affect child development.

End Notes:

1. For the NLSCY, an economic family is defined as a family in which all members are related by blood, marriage, common-law relationship or adoption; foster children are considered to be part of a family.
2. NLSCY data will be used to estimate the proportion of children having specific characteristics, for example, specific behaviours. The 'min-p.' is the 'minimum proportion' or smallest proportion obtained by the survey estimates where the estimates can be reliably released or published without qualification. If one wishes to estimate a proportion smaller than the calculated min-p, the estimates will not be reliable and should not be released. If the estimated proportion is greater than the min-p, the estimates can be released. The min-p after five cycles (ten years) was considered to be a critical criterion for the NLSCY sample because the NLSCY is a longitudinal survey. The min-p for earlier cycles should be better than the one after five cycles.
3. Because the NLSCY uses the Labour Force Survey sample frame, the LFS clusters become the basis for the NLSCY sample as well. Cluster sampling for the LFS includes, for example, the process of sampling city blocks. Only certain blocks are chosen to be part of the sample; dwellings are selected from within those blocks.