SAMPLE DESIGN ISSUES IN A LARGE-SCALE MULTIPLE FRAME NATIONAL SURVEY: THE CANADIAN COMPONENT OF THE ADULT LITERACY AND LIFE-SKILLS SURVEY

Joseph Duggan, Elisabeth Neusy, and Yves Bélanger[1]

ABSTRACT

This paper provides an overview of the design considerations in the development of the Canadian component for the Adult Literacy and Life-skills Survey (ALL). This international survey was designed to provide measures of proficiency in several literacy and life-skill domains for adult populations. The Canadian component, implemented in 2003, also profiled skill sets for targetted subpopulations such as youth, urban aboriginals, immigrants, and linguistic minorities in certain sub-national regions. Each of these provincial regions was stratified and had a two-stage (dwelling and then individual) sample for the urban portion and a three-stage sample (with geographical areas as primary sampling units) for the rural portion. A base sample of approximately 8,000 privately occupied dwellings was selected using the 2001 Canadian Census of Population and Housing as a frame for dwellings. Then, supplementary samples were selected sequentially for each region; a modified multi-frame weighting method was proposed to account for the dependencies in this design and the estimation of the variance was performed using the combined jackknife technique.

KEY WORDS: Stratified Multi-stage sample, Complex survey design, Supplementary samples.

RÉSUMÉ

Cette présentation fournit une vue d'ensemble des considérations reliées au plan de sondage dans le développement de la composante canadienne de l'Enquête sur l'alphabétisation et les compétences des adultes (EAC). Cette enquête internationale a été conçue pour fournir des mesures de compétence dans plusieurs domaines touchant l'alphabétisation et les compétences de la population adulte. La composante canadienne de l'enquête, implantée en 2003, cherche également à profiler ces ensembles de compétence pour des sous-populations telles que les jeunes, les autochtones urbains, les immigrants et les minorités linguistiques dans certaines régions infra-nationales. Chacune de ces régions provinciales a été stratifiée et a été représentée par un échantillon à deux degrés (logement, et ensuite individu) pour la partie urbaine et un échantillon à trois degrés (avec les secteurs géographiques comme unité primaire d'échantillonnage) pour la partie rurale. Un échantillon de base d'environ 8 000 logements résidentiels a été choisi en utilisant le Recensement canadien de la population et des logements de 2001 comme base de sondage de logements. Ensuite, des échantillons supplémentaires ont été choisis séquentiellement pour chaque région. Une méthode de pondération pour les bases multiples modifiée a été proposée pour tenir compte de la dépendance dans ce plan et l'estimation de la variance a été effectuée par la technique du jackknife dit combiné.

MOTS CLÉS : Échantillonnage stratifié à plusieurs degrés; Plan d'enquête complexe; Échantillons supplémentaires.

1. INTRODUCTION

The International Adult Literacy and Life-skills Survey (ALL) can be viewed as the latest in a series of surveys with the objective of measuring the functional literacy levels of the adult population. This current version, the ALL survey of 2003, was preceded in Canada by the International Adult Literacy Survey in 1994 and the 1989 Statistics Canada survey on Literacy Skills Use in Daily Activities. The international aspect of ALL is reflected in the number of countries that have participated in the development of the survey and in the implementation of their own national components of the survey. The value of ALL is based on the accepted standards and the consistency of methods used by the various countries participating in the survey. There is a team of survey professionals at Statistics Canada responsible for coordinating the international effort for ALL. And, as in every other country participating, there is also a team from Statistics Canada that

[1] Joseph Duggan(joseph.duggan@statcan.ca), Elisabeth Neusy and Yves Bélanger, Statistics Canada, 2300 Main Building, Tunney's Pasture, Ottawa, ON K1A 0T6,

is responsible for conducting the national component of the survey. This paper concentrates on the Canadian component (known as the International Adult Literacy and Skills Survey) and the elements of its design.

2. SURVEY CONTENT

The elements of functional literacy and life-skills in ALL are evaluated through measures of proficiency in Prose Literacy, Document Literacy, Numeracy, and Problem Solving. A few brief examples help to distinguish between these related and similar-sounding notions. While prose literacy can be evaluated using questions about continuous discourse such as instructions in a bicycle owner's manual, document literacy requires that information be extracted from charts, graphs, maps, and other similar printed materials. Numeracy skills are demonstrated in correctly completing the additions on a bank deposit slip, whereas problem solving could focus on successfully identifying and ordering the necessary steps in an apartment search. Every question, or set of related questions, is based on an item. These items are grouped into eight blocks: four literacy (L1-L4), two numeracy (N1-N2), and two problem solving blocks (PS1-PS2). The blocks are combined in pairs using a block incomplete booklet (BIB) design to arrive at 28 booklets.

The booklets were distributed amongst the sample according to the design for the entire Canadian sample, over and beyond the international minimum requirement of 5,400 respondents from each language tested. As each booklet can take upwards of an hour to administer, each respondent was only asked to complete one booklet. The method of spreading the blocks across booklets substantially reduced the burden on respondents, but will make the final proficiency measures unreliable for any given individual due to the fact that no one person was tested on more than just a subset of all domains. Item Response Theory will be used to obtain estimates of proficiency level through five sets of plausible scores (ranging from 0-500) on each domain for aggregations of sampled respondents.

3. SURVEY INSTRUMENTS

The main task booklets were the last of a series of collection instruments to be applied. Initially, respondents were asked to complete a survey entry component, or screener, which constructed a roster for each sampled dwelling. This screener collected enough demographic data to identify target subpopulations for the survey and to permit the random selection of one member from each dwelling. The background questionnaire was then asked of the selected respondent, encompassing several modules of information required to relate the tested skills to individuals' economic and social situations. There were comprehensive questionnaire modules dealing with labour force activities, participation in education and learning, social capital, and well-being, among others. As a result, the background questionnaire required a median time of about 35 minutes to administer. The third instrument was a relatively short core task booklet of six very basic questions. Respondents who failed to achieve a specified number of correct responses to the core task booklet were not required to attempt the increasingly more difficult main task booklet. All respondents were to attempt the core, and then if indicated, the main task booklet (median completed time of 58 minutes) immediately after completing the background questionnaire in order to control the impact of fatigue on the assessment tools.

The core and main task booklets were paper and pencil assessments; however the screener, background questionnaire, and even the administration of the core and the main task booklets were handled in a computer-assisted personal interview (CAPI) environment. As a benefit of an extensive match of the census frame to the central Address Register, telephone numbers were available for approximately 74% of the survey file. In such a case, interviewers were permitted to make an initial contact by telephone to complete the screener and to then schedule an appointment for a personal interview with the selected respondent.

4. SURVEY DESIGN

4.1 Population and Coverage

The target population is comprised of all Canadian residents that were 16 years of age or older at the time of data collection, excluding long-term institutional residents, members of the armed forces, and individuals living on Indian Reserves. The population that was in fact covered by the survey differed in a number of practical respects, but all of these exclusions combined are still within the survey standard of no more than 2% of the total population. Residents of sparsely populated regions were excluded from the survey population. It is estimated that the coverage for the survey was 98.5% nationally, with provincial coverages ranging from 95% to nearly 100%. In the northern territories, reduced levels of coverage (70-90%) were obtained because only the communities covered in the national Labour Force Survey were included (personal interviewing outside of these centres was considered either too costly or impractical).

4.2 Frame

The most recent Census of Population and Housing, with a reference date of May 15th 2001, was chosen as the frame for the survey. This already existing frame offered the ability to use reported household-level characteristics to identify dwellings with greater probability of containing an individual belonging to specific target subpopulations of interest. This auxiliary information greatly assisted the efficiency of the sample design. Specifically, the survey frame consisted of households enumerated by the Census long-form (20%) sample. The survey's national base sample, provincial top-up samples to the base, and supplementary samples related to age could have been selected from short-form households, but the long form data was required to identify the remainder of the special subpopulations. In the case of minority language samples, the quality of the long form responses is judged to be superior to that of the short form. The presence of questions on the knowledge of, and the use of languages, in addition to the mother tongue (language first learned and still understood) provide respondents with more opportunities to properly characterize their linguistic profile.

4.3 Required Precision

Three basic levels of sample sizes were proposed for the survey, a minimum, an average, and a maximum strategy depending on the precision required from each estimate. The minimum strategy corresponded to the ability to estimate a proportion (such as the proportion of the population at a given level of proficiency for a given domain) as low as 15% with a coefficient of variation of at most 16.5%, i.e., of acceptable quality. In the 1994 IALS, the smallest proportions of Canadian adults in any published skill level were found in the lowest level of proficiency; 16.6% were classed in Level 1 for Prose Literacy and a similar amount, 16.9% for Quantitative Literacy (numeracy). The Problem Solving domain was not part of IALS and so a good national estimate of this proportion will only be known once the estimates from ALL are compiled. The average strategy is very similar to the minimum, except that the ability to estimate a proportion as small as 10% was sought. Finally, the maximum strategy corresponded to the ability to produce estimates for proportions as small as 25% *within* a literacy level, provided that level represents at least 15% of the total population.

4.4 Stratification

There were two main strata constructed in every province, an urban stratum and a rural (including small urban centres) stratum. The urban stratum was restricted to urban centers of a particular size, as determined from the previous census. The remainder of the survey frame, essentially composed of Census geography, was delineated into primary sampling units (PSUs) by Statistics Canada's Generalised Area Delineation System (GArDS). The PSUs were created to contain a sufficient population in terms of the number of dwellings within a limited area of reasonable compactness. In addition, a general indication of the education level of the population from the 1996 Census was used to create PSUs that reflected (as much as possible) the educational distribution of their province.

A second, implicit, stratification was used in the systematic selection of households for each sample. The highest level of education for each adult in the household, as recorded in the Census frame, was used to determine a representation of the dominant class from four broad levels: 1) less than high school, 2) high school graduate or some post-secondary education, 3) college graduate, and 4) university graduate. Formal educational attainment is not the only, but is the main, determinant of performance in evaluations of literacy (OECD 2000). Ordering the households by education within geographic regions before sample selection increased the ability to represent a range of educational backgrounds.

4.5 Allocation

The sample was allocated between strata under a Neyman allocation, incorporating a conservative design effect of 2 for the rural stratum and 1.5 for the urban stratum. After allocation, it became apparent that several PSUs in the rural strata were sufficiently important that they were effectively being sampled with certainty. These PSUs were converted to a new pseudo-urban stratum, to be treated similar to the urban stratum in terms of sample selection.

As a final step before sample selection, the negotiated sample sizes were inflated to account for an international target minimum response rate of 70% and for mobility in terms of the characteristics of interest for each subpopulation covered by a supplementary sample. A blended rate was calculated using reported 1-year and 5-year mobility variables from the Census as proxies, and applied to the time lag between the Census and the start of collection in March of 2003. These

rates were adjusted downward in each stratum to reflect the expected replacement of movers by others with the same target characteristics for each supplementary sample.

4.6 Stages of Sample Selection

Within the urban stratum, two stages of sampling were used. In the first stage, households were selected systematically with probability proportional to size (in terms of the number of adults, capped at four for the base and at three for supplementary samples). During the second stage, the CAPI application automatically selected an individual (by SRS of eligible adults) from the household based on demographic information in the screener. Three stages were used to select the base and supplementary samples in the rural stratum. In the first stage, PSUs were selected by PPS as measured by the total number of adults for each sample's survey population from the 2001 Census. The second and third stages for the rural stratum repeated the same methodology employed in the two stages of selection for the urban stratum.

4.7 Dependent Selection of Samples

Each province had a base sample which covered the general population. Additionally, provincial ministries and other organisations sponsored supplementary samples to increase the base or to target specific subpopulations. Table 1 shows the expected number of respondents in each sample: the base, youth (ages 16-24 in Québec and 16-29 in British Columbia), adults aged 25-64 in Québec, linguistic minorities (English in Québec and French elsewhere), recent and established immigrants, urban aboriginals, and residents (specifically Inuit and non-Inuit for Nunavut) of the territories.

Table 1 – Disposition of Base and Supplementary Samples

Region	Base	Youth	Adult	Lang.	Immigr.	Aborig.	Non-aborig.	Total
Newfoundland and Labrador	1,350							1,350
Prince Edward Island	650							650
Nova Scotia	1,350							1,350
New Brunswick	650			760				1,410
Québec	1,110	815	1,885	570	270			4,650
Ontario	1,690			3,000	1,060			5,750
Manitoba	1,350			450		700		2,500
Saskatchewan	650					700		1,350
Alberta	1,350				70			1,420
British Columbia	1,350	490			280			2,120
Yukon Territory						700	700	1,400
Northwest Territories						450	450	900
Nunavut						700	180	880
Canada (Total)	11,500	1,305	1,885	4,780	1,680	3,240	1,340	25,730

After adjusting for non-response and the anticipated mobility of the target subpopulations, an overall sample size of over 40,000 was achieved. The samples were selected sequentially, one after another, starting with the base sample. After the selection of each sample, chosen households were removed from the frame before the next selections, thereby making the samples dependent. The sequential selection of multiple samples in a province can be viewed as multiple phase sampling.

To expand on this, consider the example of Manitoba, where three samples were selected: a base sample, followed by a linguistic minority (Francophone) sample and finally an aboriginal sample. The design of each sample was based on the previous samples. The base sample had a two phase design: the first phase was the 2001 census long form sample, and the second phase was the systematic selection of dwellings to create the base sample. The linguistic minority sample had a three phase design: the first phase was the census long form sample, the second phase was composed of households not selected in the base, and the third phase was the linguistic minority sample. The aboriginal sample had a four phase design: the first phase was the census sample, the second phase was the households not in the base, the third phase was the

households not in the linguistic sample, and the fourth phase was the aboriginal sample. Non-francophone aboriginal households were considered to have a zero probability of inclusion in the linguistic sample.

4.8 Probabilities of Selection

In selecting sample A, or S_A, an individual *i* was selected from household *h* of size H. The probability of selecting *h* in the Census long form sample was denoted as $P(h \in S_C)$, and the probability of selecting *h* in sample S_A, conditional on S_C, S_1, ..., S_{A-1} was denoted as $P(h \in S_A \mid S_C, S_1, ..., S_{A-1})$. The conditional probability of selecting *i* was then:

$$\begin{aligned} P_i^{(A)} &= \left[\text{Probability of selecting } h \text{ in } S_C\right] \times \left[\text{Probability of not selecting } h \text{ in } S_1, ..., S_{A-1}\right] \\ &\times \left[\text{Probability of selecting } h \text{ in } S_A\right] \times \left[\text{Probability of selecting } i \text{ from } h\right] \qquad (1) \\ &= \frac{P(h \in S_C)\left[\prod_{a=1}^{A-1}(1 - P(h \in S_a \mid S_C, S_1, ..., S_{a-1}))\right] P(h \in S_A \mid S_C, S_1, ..., S_{A-1})}{H}. \end{aligned}$$

The design weights were defined as the inverse of the probabilities of selection. An unbiased conditional estimator of Y (Särndal 1992, p. 347) for the subpopulation covered by sample A is given by: $\hat{Y}^{(A)} = \sum_{i \in S_A} \left(P_i^{(A)}\right)^{-1} y_i$.

5. WEIGHTING AND ESTIMATION

5.1 Non-response Adjustments

To begin, sample units were categorized as either respondents, out-of-scope households, non-respondent households (those without data from the screener), and non-respondent individuals (screener completed, but no data for the selected respondent). The CHAID algorithm in Knowledge-Seeker software was used successively to form weighting classes (response homogeneous groups) to adjust for non-respondent households and non-responding persons in two separate stages for each province and sample type. Afterward, the person weights of the respondents were adjusted by the factors calculated from each step in order to represent all households.

5.2 Combining the Samples

With the overlap in coverage from the various samples, it was necessary to integrate the weights to be able to produce estimates using all units from all samples. The situation is analogous to that of 'sampling on two occasions' as described in the literature (e.g., Cochran p.346, Särndal p.368), where the sample of the second occasion is composed of two sub-samples: one from the first sample and a second from the units not selected in the first sample. Both sub-samples cover the entire population. It is also comparable to a multiple frame situation, except that here the samples are dependent. The chosen solution was to partition the entire sample according to the subpopulations targetted in the supplementary samples, by defining indicator variables for each target subpopulation. Each indicator variable was assigned a value of 1 if the sampled unit belonged to the target subpopulation, and 0 otherwise. The collective sample was divided into *D* partitions according to the combinations of the observed indicator variables. A domain estimate $\hat{Y}_d^{(a)}$ was created for each sample *a* which had at least one sampled unit in the segment of the population represented by partition *d*. Where there are A samples to consider, a composite estimator, $\hat{Y}_d$, for partition *d* can be given by:

$$\hat{Y}_d = \sum_{a=1}^{A} \alpha_d^{(a)} \hat{Y}_d^{(a)}, \quad \text{where } \sum_{a=1}^{A} \alpha_d^{(a)} = 1, \ \alpha_d^{(a)} \geq 0, \text{ and } \alpha_d^{(a)} = 0 \text{ if } S_a \cap d = \phi, \qquad (2)$$

and the estimator of the population total is given by: $\hat{Y} = \sum_{d=1}^{D} \hat{Y}_d$. This method of integrating the weights was proposed by Hartley (1962) for multiple frames, and was also proposed for 'sampling on two occasions' designs. The scaling coefficients, $\alpha_d^{(a)}$, are optimal when they minimize the variance. As this involves calculating a separate set of coefficients for each variable of interest, the coefficients were instead made proportional to the realized sample sizes of the various samples within the partition. This solution is reasonable, but not optimal as it assumes equal design effects between independent samples.

5.3 Calibration

The samples were calibrated separately in each province/territory using twelve categories of combined age group and gender, two gender groups for each target subpopulations present, and 1-4 groups formed from the populations of larger Census Metropolitan Areas (where applicable). Attempts to include household size and education variables proved unsatisfactory and were abandoned. Variables that were used had been validated through matches of the collected survey data with available frame information. Small amounts of missing data for the calibration variables were imputed. Census counts for all calibration variables at the enumeration area level were inflated according to the growth measured between provincial age and gender totals from the Census and the corresponding official demographic counts as of June 21, 2003. This reference date represented an approximation of the midpoint of collection both in terms of calendar days, and in terms of completed response.

6. VARIANCE ESTIMATION

To simplify variance estimation, it was assumed that the various samples were selected independently. It is believed that this assumption causes a slight overestimation of the variance. The assumption allowed the use of jackknife variance estimation as proposed by Lohr and Rao (1997) for multiple frames. They propose treating the samples from the different frames as samples from different strata, and apply the jackknife as for stratified sampling. To meet international standards, 30 jackknife replicate weights that cut across strata will be included on the survey's micro-data file.

REFERENCES

Cochran, W.G. (1977). *Sampling Techniques* (3rd. ed.). New York: John Wiley.

Lohr, S. and Rao, J.N.K. (1997). Jackknife Variance Estimation in Multiple Frame Surveys. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, pp. 552-557.

Hartley, H.O. (1962). Multiple Frame Surveys. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, pp. 203-206.

Organisation for Economic Co-operation and Development, OECD (2000). *Literacy in the Information Age: Final Report on the International Adult Literacy Survey*. Ottawa: Statistics Canada.

Särndal, C.-E., Swenson, B., and Wretman, J.H. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Special Surveys Division (1999). *International Adult Literacy and Skills Survey: National Planning Report*. Ottawa: Statistics Canada.