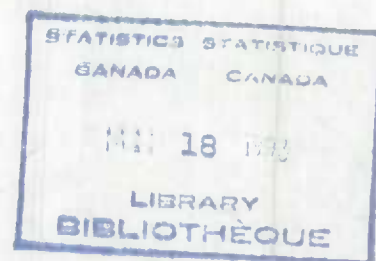


COMPUTER ASSISTED INTERVIEWING DATA QUALITY TEST

Ritu Kaushal and Normand Laniel, Statistics Canada



ABSTRACT

A test was conducted for the Canadian Labour Force Survey (LFS) to determine if a change from Paper and Pencil Interviewing (PAPI) to Computer Assisted Interviewing (CAI) would produce a break in the labour force series and affect the quality of data. The change in data quality may also have an impact on the labour force estimates. The design of the 21 week test used two groups of interviewers, a control group using PAPI and a test group using CAI. Each test assignment was paired with a control assignment, so as to diminish differential effects due to geography. For part of the test, both groups used PAPI. This feature was used to estimate systematic differences between the control and test groups. Based on this test, no statistically significant differences were found in the labour force estimates between the two groups. Of the data quality measures, the analysis on household non-response rate was inconclusive but the vacancy rate was larger for CAI than PAPI. Based on the positive results from this test, CAI is being implemented for the LFS.

KEYWORDS

Mode Effect, Series Break, Non-response, Vacancy rate

1. INTRODUCTION

The Canadian Labour Force Survey (LFS) is undergoing a redesign which encompasses all aspects of the survey. One of the proposals considered for the data collection process is a change from paper and pencil interviewing (PAPI) to computer assisted interviewing (CAI). A series of tests has been conducted to evaluate this proposal. The test described here was used to determine if a change in collection mode had an impact on the quality of data collected and, more importantly, whether it had an effect on the labour force series. This paper deals with the effect of change in mode on the three LFS rates: Employment to Population ratio, Unemployment rate and the Participation rate. Non-response rates and vacancy rates are the only quality indicators studied here.

Computer Assisted Interviewing (CAI) is a means to compress interviewing, data capture and preliminary editing stages into one integrated step. Interviewers are equipped with a notebook computer and an electronic questionnaire as opposed to the paper forms used in PAPI. The information is captured electronically during the interview. The three anticipated advantages of CAI are reduced cost and time, and better quality of data. First, there is a substantial reduction in the on-going cost of conducting a survey because of savings in data entry resources. Second, the data becomes available more rapidly because capture is immediate and the data can be transmitted electronically rather than by shipping,

batching and capturing forms. Last, an improvement in quality can be expected because the computer controls the sequence of questions. CAI also applies range and validity checks, and edits while entering data and the respondent is present to resolve any discrepancies.

The LFS is a monthly household sample survey of 59,000 households, that produces the official unemployment rate and other important labour market statistics in Canada. A change in the labour force series due to operational or methodological changes is undesirable for two reasons. First, any change of the series is likely to be misinterpreted as a fluctuation in the labour market rather than a function of the measurement instrument. Second, an interruption of a time series affects seasonal adjustments made in the future. Therefore, the impact of a change in collection methodology must be evaluated. To get an unbiased measure of such a change, it is necessary to isolate the effect of the data collection instrument from other effects due to time (natural changes in the labour market), interviewer, type of contact (personal or telephone), geography, etc. The test is designed to separate the instrument effect from these other effects and thereby provide an accurate comparison between CAI and PAPI. This is discussed further in Section 2.

Currently, in the LFS, data is collected through eight Regional Offices (RO) across Canada. The regional offices receive computer files from Head Office (HO) specifying the sample selected. The geographic information about the sample is used to plan interviewer assignments. Regional offices print pre-filled forms with geographic identifiers for all dwellings. For dwellings entering the sample, this is the only pre-filled information sent to interviewers. These birth dwellings constitute a sixth of the sample. For the remaining five-sixths (the non-birth dwellings), demographic and some job description information, from the previous month, is pre-filled on the forms. These pre-filled forms are packaged into an assignment, along with extra blank forms, letters of introduction, maps, listings and control forms, and sent to interviewers. For CAI, the information for the pre-filled forms is transmitted electronically.

Under the current design of the LFS, sampled dwellings stay in the sample for 6 months. LFS interviews are conducted during the week following the reference week each month. A typical assignment contains 60 dwellings. Generally, the birth interviews are conducted in person at the dwelling and subsequent interviews are done over the telephone. Proxy interviews are accepted from household members 15 years of age or older. At the end of the day, PAPI interviewers ship completed questionnaires to the ROs.

At the RO, the clerical staff perform range and validity checks on the PAPI forms received. The forms are separated into births and non-births and sent for data capture in batches. The captured data undergoes validation edits and sample verification before being transmitted to the HO for further processing. At the HO, the data received from the RO are edited and coded.

There are some differences between CAI and PAPI. CAI interviewers enter the respondents' answers directly into the computer. The interviewer is forced to enter the questions in the given sequence and follow the skip pattern. Out of sequence questions and shortcuts through the questionnaire are not permitted. Therefore, in CAI, the questionnaire is applied more rigorously and uniformly, which is expected to provide cleaner data. Some edits are programmed into the software that supports the questionnaire. Edit failure messages appear on the screen if there is any conflicting information entered. When the interviewer is with the respondent, she is able to reconcile the information immediately. It should be noted that verification of keying errors is not possible for CAI in this test. When a case is complete, the CAI interviewer no longer has access to the case. Due to the above mentioned differences

and the extra time for keying of data and running edits, the CAI interview is expected to take longer than the PAPI. Thus, the response burden is expected to increase, which may in turn affect non-response.

2. DESIGN OF THE TEST

The objective of the test was to determine whether the impact of CAI on the unemployment rate would exceed one percentage point. The sample size was based on this criterion. Other labour force rates and data quality were also monitored throughout the test.

This test was conducted independently of the regular LFS but using trained LFS interviewers. Four ROs were selected on the basis of operational considerations. Primary Sampling Units (PSUs) were chosen arbitrarily (not randomly) in some strata within the geographic region in the jurisdiction of each RO. Therefore, the sample was not representative of either provinces or Canada. It was representative only of the selected strata that were sampled. The PSUs were chosen from those already listed, which had unused sample, and were based on the availability of LFS trained interviewers.

The design selected used a control group of PAPI interviewers and a treatment group of CAI interviewers. The budgetary constraint of number of computers and the operational constraint of not mixing CAI and PAPI assignments for an interviewer were the main reasons for this selection. Forty two interviewers were used for the test. Half of them were trained to use CAI. Test and control interviewers were paired and each pair covered the same geographic area. Paired assignments were independent samples from the same population of dwellings.

Interviewers in a pair were "randomly"¹ assigned to the control or test group. The test group of CAI interviewers covered 50 dwellings each week (total CAI 1050 per week) and the other half of the interviewers' assignments included 50 dwellings per week (total PAPI 1050 per week). In November, both sets of interviewers conducted PAPI interviews. The design can be summarized by Table 2.1. Here, the shaded areas represent the PAPI interviews (and PAPI interviewers) and the blanks represent the CAI interviews. The rows represent two different assignments that have been paired by design. The columns are the months and weeks of the survey.

Table 2.1: Use of Instrument by Control and Treatment Groups over 21 Weeks

Weeks →	0a 0b 0c			1a 1b 1c			2a 2b 2c			3a 3b 3c			4a 4b 4c			5a 5b 5c			6a 6b 6c		
Assignment ↓																					
Control																					
Treatment	C	A	I	C	A	I	C	A	I	C	A	I				C	A	I	C	A	I

The test was conducted for 21 consecutive non-LFS weeks over a seven month period. For each test

¹ Operational constraints may have limited the ability to satisfy this requirement.

week in a month, an interviewer had a different assignment of dwellings selected from a different set of PSUs. Therefore, each interviewer had three assignments in a month. Each assignment followed the LFS rotation pattern after the phase-in was completed. For this test, three rotations were phased-in at the first month of the test. The other three were birthed in the second month. After this phase-in period, the LFS rotation pattern was followed. The following table describes the rotations, replacement, and size of the sample in this test.

Table 2.2: Number of Households to Interview for Control Group

Rotation	Month	July (0)	August (1)	Sept (2)	Oct (3)	Nov (4)	Dec (5)	Jan (6)
3	phase-in	525	525					
4	phase-in		525	525				
5	phase-in	525	525	525	525			
6	phase-in		525	525	525	525		
1	phase-in	525	525	525	525	525	525	
2	phase-in		525	525	525	525	525	525
3	replacement			525	525	525	525	525
4	replacement				525	525	525	525
5	replacement					525	525	525
6	replacement						525	525
1	replacement							525
Total		1575	3150	3150	3150	3150	3150	3150

Both groups used PAPI in November so that the mode effect could be isolated from any systematic differences between the control and treatment groups. November could then be used as a benchmark to adjust the other months for systematic effect.

3. RESULTS

The impact of the mode of interviewing on the labour force rates, unit non-response and vacancy rates is analyzed here. July was excluded from the analysis as there were software problems that may have affected the labour force status and the quality measures. These problems were rectified and the data from August to December was used. January data was not available for this paper. All the differences between test and control rates are expressed as the control group rate subtracted from the test group rate.

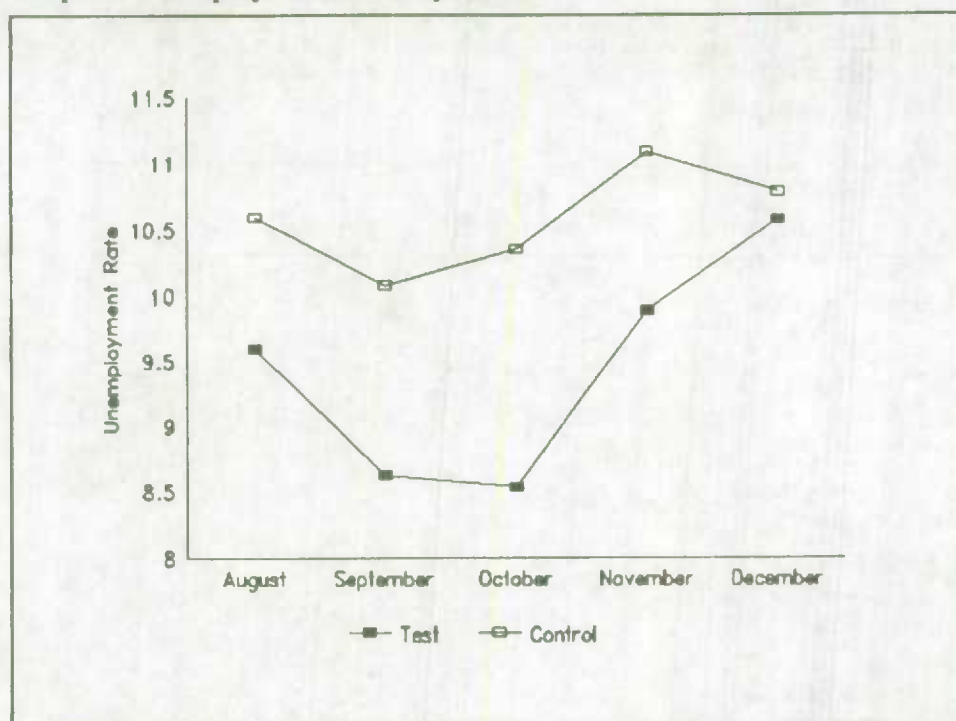
3.1 Labour Force Rates

The three labour force rates discussed in this paper are the unemployment rate, the participation rate and the employment to population ratio. Based on responses to the questionnaire, respondents are classified as being in or out of the labour force. The respondents in the labour force are further categorized as being employed or unemployed. The unemployment rate is the percentage of the

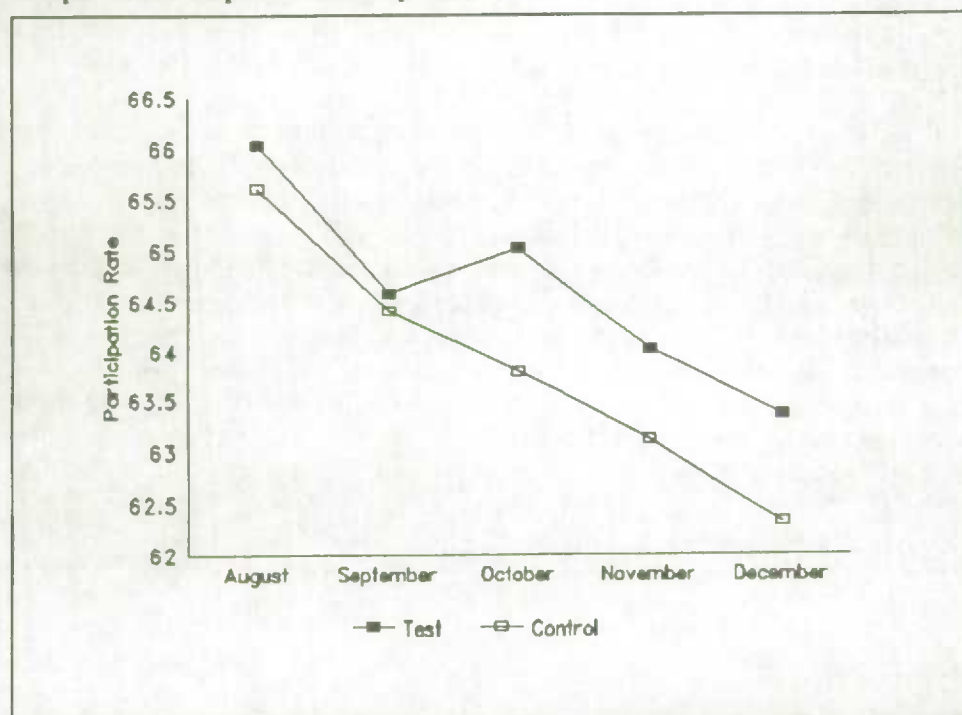
unemployed to those in the labour force. The participation rate is the percentage of the labour force to the population. These three rates are correlated. For example, with the participation rate held constant, if the unemployment rate goes up then the employment to population ratio decreases.

Graphs 1 to 3 show the three major labour force rates for test and control groups. For the unemployment rate in Graph 1, there appear to be appreciable differences between the rates for all months except December. The differences from September to November are statistically significant at a 95% confidence level, while the difference in August is marginally significant. This trend of differences appears to persist despite the sampling variability introduced by sample rotation and the use of PAPI by both groups in November. This suggests that a large portion of the differences observed can be attributed to factors other than mode. The size of the mode effect may not be large enough to be detected by the test. In December the difference appears to be much smaller. There were two changes in December, namely sample rotation and change of mode from PAPI to CAI for the test group. It is possible that rotation brought the two rates closer together.

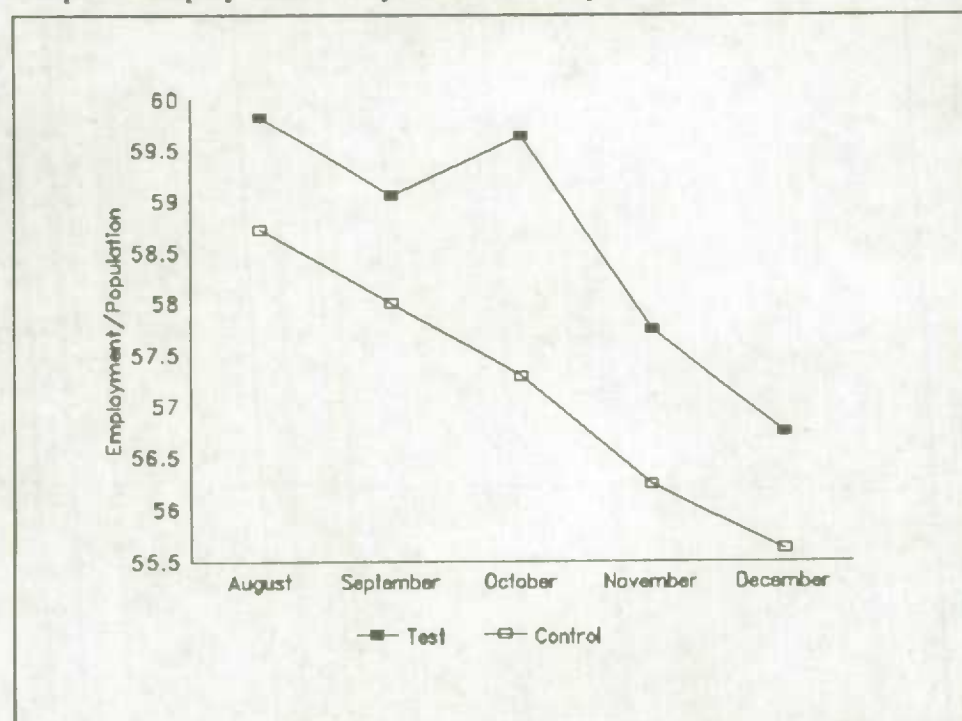
Graph 1: Unemployment Rates by Month



Graph 2: Participation Rate by Month



Graph 3: Employment to Population Ratio by Month



None of the differences in the participation rate seen in Graph 2, are statistically significant. However, this does not necessarily imply that there is not a mode and/or systematic difference between the test and control groups. The sample sizes for this cannot detect monthly differences of less than about 1.6 percentage points for the participation with 95% confidence.

In Graph 3, the difference between the two employment to population ratios is approximately constant except in October. Only the difference in October is statistically significant at a 95% confidence level. Other months are marginally significant with November being significant at a 90% confidence level. The overall trend suggests differences between the two groups. Because there is no noticeable change from this trend in November, it is concluded that a large part of this difference was due to systematic differences not related to the mode.

To see whether these differences in labour force rates may be attributed to the mode effect, the November data was investigated because both sets of interviewers used the PAPI instrument. The following table presents the differences between control and test groups in November:

Table 3.1: Systematic Differences in November

Regional Office	Estimated Effect	Standard Deviation	P value
Unemployment Rate	-1.20	0.55	0.03
Participation Rate	0.89	0.86	0.31
Employment/Population	1.51	0.89	0.09

As indicated in Table 3.1, systematic differences persist even when both groups use the same instrument. There are several possible reasons for these systematic differences between the control and treatment groups. First, interviewer characteristics that impact on the LFS rates may not have balanced out between the two groups. Some examples of interviewer characteristics are length of LFS experience, LFS training, understanding of LFS concepts and interviewing skills. Second, there may be operational differences between the two groups, such as supervision during the test. Third, there may be systematic differences in the pairing of assignments. Because of the smaller population in rural areas, the pairing of assignments was done at the PSU level (ie. a group of clusters). As a result the NSR assignments comprised of different clusters. Conversely, in urban areas, the pairing was done at the cluster (ie. city block) level.

Depending on the level of confidence chosen, the effect on unemployment rate ($p=.03$) and the employment to population ratio ($p=.09$) may be taken to significant. The level of significance is a function of the sampling variability and not necessarily an indication of the absence of systematic or mode effect. Therefore, to estimate the mode effect, an adjustment for systematic effect was made even if the test did not show this effect to be significant for all the rates.

The model used to estimate the mode effect for the labour force rates is described in detail in the Appendix. The model was formulated to reflect variables influencing the LFS rates. Then the extraneous effects were removed by differencing within pairs and between months. For example, in a given month

j (not including the benchmark November, month=4) and week k, for a given area l, interviewers are distinguished by value of i, which indicates which group the assignment belongs to. The control group is indicated by a C and the test group by T. If all other effects were controlled for, then the effect would be defined as the difference in rates between the CAI and PAPI groups (the quantity in the first set of parentheses in equation (1)). To take the systematic effect into account, the difference in November (in the second parentheses) was subtracted from the difference in any given month. The following difference of rates represents an estimate for the mode effect for a given week, month and interviewer pair:

$$effect_{jkl} = (rate_{Tjkl} - rate_{Cjkl}) - (rate_{T4kl} - rate_{C4kl}) \quad (1)$$

These differences were obtained for each pair of interviewers for a given month and week. The mean of these observations gave an estimate of the mode effect from the test. The standard deviation of the effect was determined by calculating the sampling variance associated with this mean (see Appendix).

The rates used in this study were adjusted population and labour force counts. These age-sex-non-response adjustments were made to the CAI counts, based on those observed in the PAPI sample population. The following table uses the above model and is based on data from August to December.

Table 3.2: CAI Effect on the Labour Force Rates

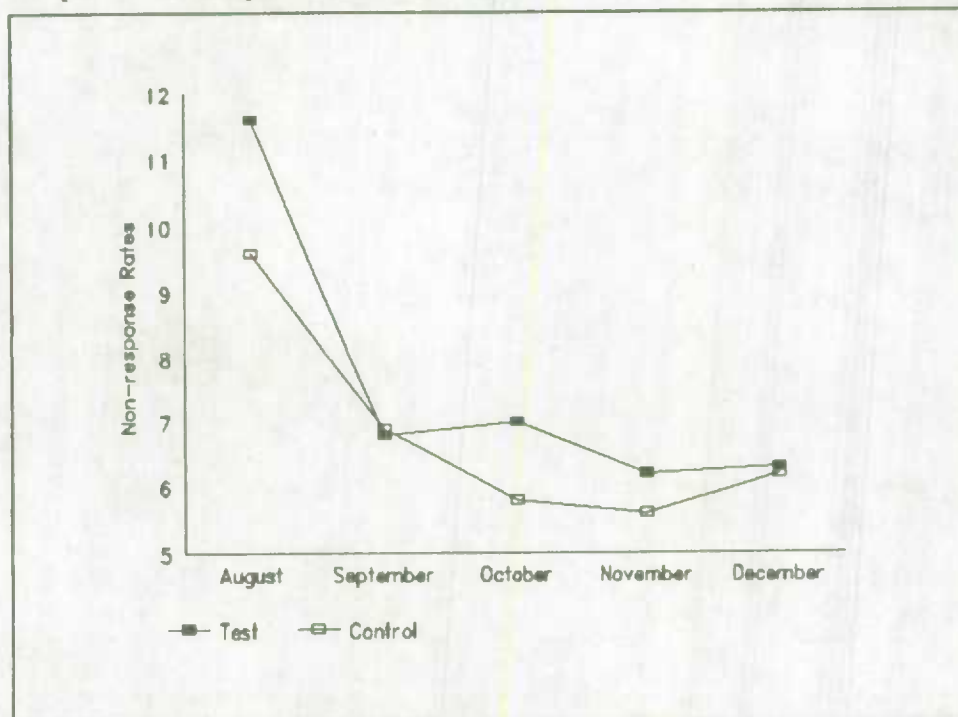
Regional Office	Estimated Effect	Standard Deviation	P value
Unemployment Rate	0.03	0.42	0.94
Participation Rate	-0.04	0.60	0.95
Employment/Population	0.07	0.62	0.91

From this table, the p-values do not suggest a mode effect. Although the point estimates are close to zero, the interval estimates are wide. For example, the effect on the unemployment rate is estimated to be between -0.81 and +0.87 for a 95% confidence level. Therefore, the point estimates should be interpreted with caution.

3.2 Household Non-Response Rates

The household non-response rate is defined as that portion of the occupied dwellings in the sample for which no interviews were completed with any eligible members of the household. In a previous CAI test

Graph 4: Non-Response Rates by Month



conducted for the LFS, Gambino et al. (1993) indicated a higher initial non-response rate for CAI initially which levelled off as interviewers became more efficient in the use of CAI.

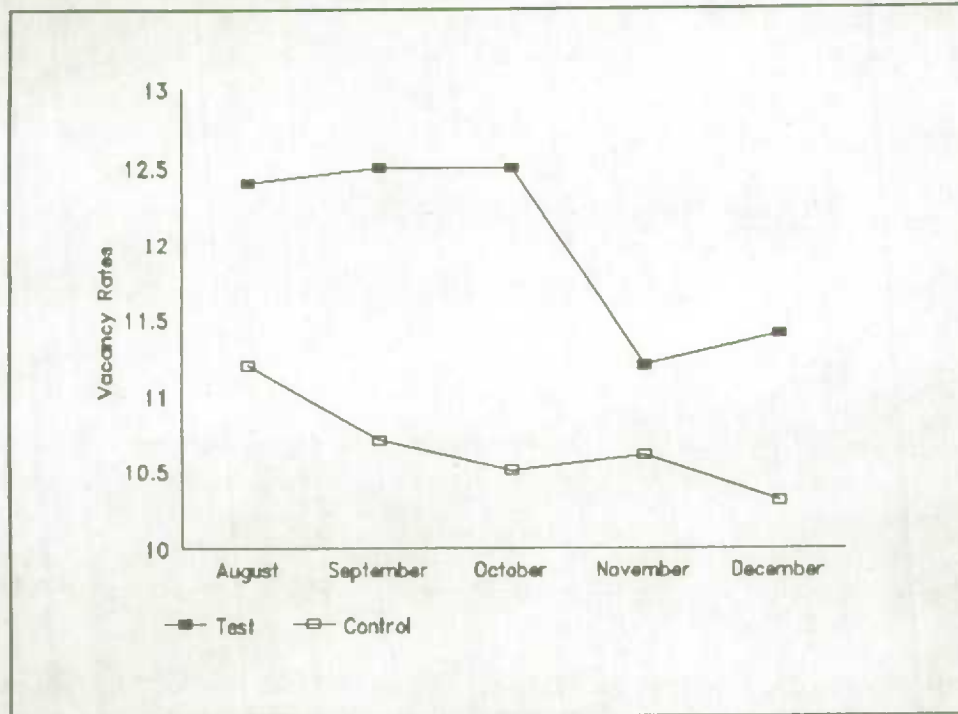
Graph 4 gives the non-response rates (expressed as percentages) for the test and control groups. The difference in August is statistically significant. In August the sample was being phased-in (see table 2.2) and the workload was high for both groups, due to the larger proportion of personal interviews. In August due to operational reasons CAI interviewers could not perform as many follow-ups as the PAPI group. For the remaining months, every effort was made to ensure that the same number of follow-ups were carried out for both the groups. Therefore, part of the differences can be attributed to operational differences between the two modes, not the mode itself. September to December do not indicate a significant mode effect.

After adjusting for a systematic effect with the November data, the overall mode effect was estimated to be 0.20 percentage points with a standard deviation of 0.36. August was not included in the estimation of this effect because of the aforementioned reasons. Based on the test, no significant mode effect was detected. From this test, it was not possible to draw any conclusions regarding the existence of a learning curve and its effect on non-response, since the July and August data were not representative of implementation of CAI due to operational constraints.

3.2 Dwelling Vacancy Rates

A dwelling is defined to be vacant if it is unoccupied, under construction, non-existent or occupied by out of scope individuals. The vacancy rate is the ratio of vacant dwellings to dwellings in the sample, expressed as a percentage. This is a quality measure routinely monitored by the LFS because it can reflect coverage problems and it is used for controlling the size of the monthly labour force sample.

Graph 5: Vacancy Rates by Month



Graph 5 gives vacancy rates for the two groups. September and October show statistically significant differences between the test and control groups. While other months do not have statistically significant differences, the trend indicates that there are differences. Because November has the smallest difference, it is possible that the drop in November may be due to the removal of the mode effect.

After adjusting for the systematic difference in November, the mode effect was estimated to be 0.93 percentage points with a standard deviation of 0.44. Therefore, based on a 95% confidence level, there is a significant mode effect on the vacancy rates. A possible reason for the mode effect is that interviewers did not have access to a questionnaire (case) after completion and were unable to make corrections if new information became available. This suggests that CAI software should be made more flexible to allow access to completed cases.

4. CONCLUSION

From this test, we can conclude that the change from PAPI to CAI does not have a statistically significant impact on the labour force series, at a 95% confidence level. For the non-response rates no statistically significant long term mode effect was observed and the analysis regarding the presence of a learning curve was inconclusive. A mode effect on the vacancy rate was found to be statistically significant.

Further analysis on the systematic effect, cost analysis and other quality measures is being carried out. Statistics Canada has decided to implement CAI beginning in September 1993. CAI will be phased-in over seven months. The mode effect and quality indicators will be monitored during implementation. In less than a year, CAI will be fully implemented in the Canadian Labour Force Survey.

REFERENCES

GAMBINO, J., GIROUX S., KAUSHAL, R., LANIEL, N. and LINDEYER J. (1993) "LFS Computer Assisted Interviewing Data Quality Test II". Working document, Statistics Canada.

TAM, S.M. (1984). "On Covariances from Overlapping Samples". The American Statistician. Vol. 38, No. 4.

SINGH, M.P., DREW, J.D., GAMBINO J.G. and MAYDA F. Methodology of the Canadian Labour Force Survey 1984-1990. Statistics Canada publication. Catalogue number 71-526.

APPENDIX: Modelling the CAI Effect on the Labour Force Rates

The following main effects were used for the model:

Instrument: This is the effect to be measured in this test. The instruments identified by the control/test group and month. The control group (C) used PAPI and the test group (T) used CAI except in month 4, when both groups used PAPI.

Time: Time is given by two different variables, month and week. These need not be given by two different variables, however the notation is greatly simplified by splitting them up because for this test weeks represent independent samples and months represent overlapping samples.

Month: Months take on values 1,...,6, where 4 (November) is the benchmark month.

Week: The values are A,B,C for the three weeks of tests in a month.

Area: (21 Levels) Different areas have populations with different socio-demographic and labour force characteristics. For this test, 21 areas were sampled. Paired assignments represent independent samples from the same area. Note that pairs correspond to areas. Different pairs interview in different areas. There may be systematic differences due to pairing at a higher geographic level in the rural (NSRU) areas.

Systematic: This is the systematic effect associated with an interviewer or assignment. This may be due to interviewer characteristics, such as experience, interviewing style, age or gender or may be external factors such as supervision or assignment. In this test, for the first 3 months the main effect due to an interviewer appears only in combination with the instrument effect. However, the switching of the CAI interviewers to a month of PAPI interviews eliminates the first order systematic effect.

The model:

$$rate_{ijkl} = \mu + \alpha_{ij} + \beta_j + \gamma_k + \zeta_l + \rho_{jl} + \delta_{il} + \epsilon_{ijkl} \quad (2)$$

where,

α_{ij} : instrument effect given by the following,

$$\alpha_{ij} = \begin{cases} \alpha_{CAI} & \text{if } i=T \text{ and } j \neq 4 \\ \alpha_{PAPI} & \text{otherwise} \end{cases}$$

i: interviewer group, C=control and T=test

μ : constant overall mean β_j : month effect ($j=1, \dots, 6$)

γ_k : week effect ($k=1, 2, 3$)

ζ_l : area effect ($l=1, 2, \dots, 21$)

ρ_{jl} : interaction between area and month due to sample rotation

δ_{il} : systematic effect for i^{th} group and l^{th} area (21 pairs)

ϵ_{ijkl} : sampling error

For months 1,2,3,5,6, subtract $rate_{i4k}$ (for benchmark month, when the CAI trained interviewers conducted PAPI interviews) from $rate_{ijk}$. Note that the instrument, i , will be different for the CAI interviewers. The area will be the same, but due to rotation, different clusters appear in different months. This change in sample occurs in both assignments in a pair. But the systematic effect, which is assumed to be constant over time and sample, will be removed. The effect of the instrument, month and area will remain. For CAI interviewers, we obtain the following:

$$rate_{Tjk} - rate_{T4k} = \alpha_{CAI} - \alpha_{PAPI} + \beta_j - \beta_4 + \rho_{ji} - \rho_{4i} + \epsilon_{Tjk} - \epsilon_{T4k} \quad (3)$$

But for the PAPI portion, the instrument does not change in month 4, therefore equation (3) for PAPI interviewers becomes:

$$rate_{Cjk} - rate_{C4k} = \beta_j - \beta_4 + \rho_{ji} - \rho_{4i} + \epsilon_{Cjk} - \epsilon_{C4k} \quad (4)$$

Subtracting equation (4) from equation (3) for the same month, j , and for test ($i=CAI$ or $PAPI$) and control ($i=PAPI$) interviewers in the same pair, represented by l .

$$\begin{aligned} d_{jkl} &= rate_{Tjk} - rate_{T4k} - rate_{Cjk} + rate_{C4k} \\ &= \alpha_{CAI} - \alpha_{PAPI} + \epsilon_{Tjk} - \epsilon_{T4k} - \epsilon_{Cjk} + \epsilon_{C4k} \end{aligned} \quad (5)$$

Therefore, with this differencing only the mode effect and a sampling error term remain. Equation (5) gives the effect for a particular pair of interviewers in a given week and month (except month 4).

The overall effect is given by the following equation:

$$\Delta = \frac{1}{JKL} \sum_{k \in \Omega} \sum_{j=1}^K \sum_{l=1}^L d_{jkl} \quad (6)$$

where, $\Omega = \{1,2,3,5,6\}$ and J is the cardinality of Ω .

Note that this estimator is equivalent to a combination of separate ratio estimators since rates are calculated at the interviewer workload level.

The variance estimator is given next. The assumptions for calculating the variances are as follows:

- (i) The population of interest is the union of the population in the clusters that were included in the sample. These clusters/primary sampling units² (PSUs) in the test were selected from a set of clusters. To be able to pair the assignments, the sampling from this set of clusters was

² Primary sampling units are sampled in Non-Self-Representing (NSR) areas.



judgmental. For this reason, the first level of sampling cannot be taken into account while calculating the variance.

- (ii) Within the clusters/PSUs, dwellings were sampled using systematic sampling. For the purposes of variance estimation it is assumed that individuals were randomly sampled (SRS) from the cluster.
- (iii) The finite sampling fraction can be ignored.

The variance is calculated for Δ as follows:

$$var \Delta = \frac{1}{(JKL)^2} \sum_{k=1}^K \sum_{l=1}^L var \sum_{j \in Q} d_{jk} \quad (7)$$

The samples from different months for the same interviewer pair are correlated because of overlapping samples. The sampling covariance between the test and control assignment sample in the same area is zero because the two samples are independent for all intents and purposes.

$$var \sum_{j \in Q} d_{jk} = var \sum_{j \in Q} (r_{Tjk} - r_{T4k}) + var \sum_{j \in Q} (r_{Cjk} - r_{C4k}) \quad (8)$$

The covariances for the test group portion of the above equation, are given below. The control group variance can be calculated in a similar fashion.

$$var \sum_{j \in Q} (r_{Tjk} - r_{T4k}) = \sum_{j \in Q} \sum_{j' \in Q} cov(r_{Tjk}, r_{Tj'k}) - cov(r_{Tjk}, r_{T4k}) - cov(r_{T4k}, r_{Tj'k}) + cov(r_{T4k}, r_{T4k}) \quad (9)$$

The estimator used for variance for rotating samples is given by Tam (1984). For a given rate, the covariance can be approximated by:

$$cov(rate_{ijk}, rate_{ij'k}) \approx C_{ijk, j'} \frac{\min(rate_{ijk}, rate_{ij'k}) - rate_{ijk} rate_{ij'k}}{\sqrt{n_{ijk} n_{ij'k}}} \quad (10)$$

where,

$C_{ijk, j'}$ = proportion of overlapping sample between months j and j' for group i , week k and area l .

n_{ijk} = sample size for $rate_{ijk}$.