Building a New Mixed Mode Cross-Platform Canadian Audience Measurement

By Don Williams (Amalco), Mark Wood (TNS), Peter Chan (TNS)

Why change now?

The Canadian media landscape has changed considerable over the past decade. Fragmentation once seen in print, radio and television, has now been replaced by consolidation of media channels. Some of the changes in the print media landscape have been as a result of economic pressures, others as a result of changing consumer behaviour combined with the increasing growth in digital.

Stakeholders from both the newspaper and magazine industries recognized the need to build a better audience measurement. Buyers and publishers agreed that the new measurement should provide a single source of compatible currencies for both magazines and newspapers. This measurement should allow capture of the entire brand footprint (print and digital media) and provide data on cross-platform use. Finally it has to be future proofed, allowing for longer term adoption of additional measurement possibilities such as a passive measure or more sophisticated sampling protocols which are projectable to the population.

After a rigorous multi-stage RFP process, a new audience measurement was finalized (with TNS as data collection supplier) in June 2014.

Previous changes

In the past, the newspaper (NADbank) and magazine (PMB) measurements ran independent of each other with different methodologies: telephone interviews for newspapers and personal interviews for magazines. The status quo of these measurements was largely a result of both methodologies providing the most reliable methods for estimating respective print readership.

Nevertheless, both measurements were attempting to move forward into the ‘brave new world’ but with a decidedly cautious approach. This included a thorough review of the use of alternative data collection methodologies - in whole or part - as well as other specific tests related to questionnaire flow or sample coverage along with experiences across the globe. All the data gathered supported a move to find a remedy for 1) declining response rates and coverage, 2) the rising costs associated with probability based sampling, and 3) the ever-changing consumer.

As a result, the newspaper measurement moved to a hybrid approach in 2012 by blending mobile phone only households (sampled from online access panels) with the RDD telephone sample. This was a significant step as it meant that, for the first time, newspaper audience data in the top Canadian markets consisted of a blend of non-probability and probability based samples. Additionally, recall measurement of a total footprint (print, web, PDF, app) also began around 2012.

PMB took a slightly different approach; the sampling methodology remained in place and instead concentrated on exploring ways to improve the quality of its data collection process and moved from pen/paper surveys to CAPI in 2012.

A new study, a new methodology

The new audience measurement methodology replaced the separate magazine and newspaper surveys with a single study that had 1) larger sample size, 2) more frequent and quicker data releases, and 3) measurement of all publications' print and digital products. Specifically,

Design

- N=36,000 completes per year from a probability based sample aged 12 years and over, sampled at a market level and built to national sample
- Respondents recruited offline using random digit dialling to take part in online survey
- Additional sampling initiatives/modes employed for better coverage of 18-24, mobile phone only households and professionals/managers (including a facility for people without online access to complete interview).

Data Collection

- Data for publications' total footprint (print, web, PDF, app) collected in a "brand-centric" mode.
- Readership data collected using Recent Reading
- Engagement data for magazine and newspapers collected using existing behavioural measures
- Product and attitudinal data previously collected was incorporated in the new single source survey - split sample approach with double ascription
- Interviewing to begin in January 2015.

Data

- Quarterly releases of all Readership and Product data with a rolling 12-month database, incorporating the most recent quarter of data into that database.

Survey pilot

Although the aim is to commence with audience measurement using the new single source platform in January 2015, a pilot phase was necessary to assess potential challenges related to the new methodology. Specifically, the two objectives of the pilot phase were to:

1. Confirm mechanics of data collection methodology; and
2. Provide information on expected audience changes to help manage transition to new data collection platform

Nevertheless, there were many challenges with implementing a new methodology within such a short period as - unlike other jurisdictions where there was considerable time between testing and implementation - the gap between from RFP acceptance to implementation was only 6 months.

Although the scope of the pilot was confined to the two largest markets (Toronto and Montreal, which also included a natural diversity by language spoken - English and French), a robust sample was still required in each market to allow for granular reviews of the data. Moreover, all of the mechanics on delivering a more inclusive sample were to be tested as well - this was intended to replicate the new method on all levels.

The specifics of the pilot were:

- Random digit dialing to recruit for online completion of readership and product survey
- Field period: October 2 to November 14, 2014
- N=2000 Readership interviews in each of Toronto and Montreal
- 10% of sample from mobile phone only households (these are initially screened via IVR)
- 5% oversample of 18 to 24 year olds (done through online access panels and river sampling)
- Oversampling on areas with high income / managerial / professionals to account for differential completion rates
- Condensed product survey of ~20 categories with expected completion rate of 60%

Overall, the test was successful mechanically as 5 of the 6 goals were met.

1.	4,000 Readership surveys	✓	Over achieved (N=4,127)

2.	2,640 Product surveys	✓	Over achieved (N=2,987 or 72%)
3.	Probability samples in each market	✓	
4.	5% 18-24 oversample	✓	(N=223 or 5%)
5.	High income / professional oversample	✓	
6.	10% Mobile phone only households		Under achieved (N=340 or 8%)

Several insights into the methodology were derived from the pilot:

- The average interview length was 40 minutes (versus an expected 45 minutes)
- Total completion target met despite lower than expected completions levels, this was improved through higher incentives
- Large volume of mobile exchanges required for mobile only household sample given the low response rates of IVR surveys in general. When combined with the truncated field period for the test, this contributed to the under-achievement of this objective as sufficient time must be built in to deliver on the total completes needed
- The sample tended to skew older than the general population but gender was more in line with the population
- Some demographic groups needed a longer fieldwork period to be sampled closer in line with population estimates.

At the readership level, there were a number of critical differences versus historical data that limited comparisons such as…

Versus previous newspaper data	**Versus previous magazine data**
1. Sample Recruitment Protocol	1. Sample Recruitment Protocol
2. Data Collection Period	2. Data Collection Period
3. Readership Methodology	3. Online Interviewing
4. Aided Logo Recall	4. Brand Footprint Measurement
5. Incentives	

With all the differences noted above it is not surprising that newspapers changes (increases) in AIR were not consistent across all newspapers but the rank order and relativity of that rank generally held in each market. That is there was …

1) little or no change in rank (e.g., $1^{st} = 1^{st}$, $6^{th} = 6^{th}$)
2) expected movements between the mid-ranked publications (that is 2^{nd} through 5^{th}) were in line with previous data, and
3) relativity held (ratio of 6th/1st) comparable to historical data.

Magazines AIR showed more stability than was experienced when Through The Book was changed to Recent Reading (average increase of 30% versus 140% increase when TTB changed to RR). Additionally,

1) Digital component adds substantially to total footprint for each group
2) 19 publications remained in virtually same rank order position as previously (compared to only 6 when RR adopted).
3) Relative ranking of titles is more stable following RR adoption (move from Through The Book)

Survey pilot summary

In conclusion…

- Survey Design - no significant issues in methodology
- Sampling - no apparent sampling Issues
- Readership results - no significant changes in ranking for magazines or newspapers but digital platform audiences do not correlate to actuals

- Product results - no indication of negative results for product data however the time to complete the condensed test interview indicated that high completion rates would not be seen.

Concern with the accurate recall of the specific digital platforms (PDF, web, or App) was noted but the discrepancy was likely a result of misattribution. Wording changes were implemented as an attempt to mitigate this in the future.

Completion levels of the product survey also was highlighted as a potential issue and ‘question load’ was reviewed and additional sections were added to allow more splitting of the sample and thus reducing respondent burden.

With these changes in place, it paved the way for the new study went to begin data collection as a national sample as scheduled in January 2015.

Expectations of ‘new’ data/challenges

The overall expectation was that readership/audience data was going to be different from historical data given all the changes, but that there should be some consistency with pilot. However, given that data collection was going to be completed in a variety of diverse geographies across Canada, it was not expected that major market trends would always hold true in less urban locations. Initial reviews of the first 6 months of data confirmed these points.

Part of the transition plan in moving from the old to the new was to smooth the introduction of the new survey and new audience results by fusing the most recent historical data (captured under the old methods) to the newly collected data to form a robust database. Fusion preserved readership levels perfectly by age, gender and geography for all magazines and newspapers (as it was controlled at these levels) but the remainder of the demographics diverged significantly from historical norms. This was especially notable for magazines given that the newspaper readership demographics were used as the input for much of the fusion. Although the transitionary database held at top-line levels, detailed readership profiles were not possible to develop with the fused data given the divergent demographics. This transition plan was shelved so that the first release would be based on 100% of new data and encompass the first 6 months of data collection (January to June) only.

Review of 6 months of data

Although the transitionary database was shelved, analyzing the first 2 quarters of data provided access to a large sample of over 19,000 completed interviews (and comparable sample sizes in Toronto and Montreal to the pilot).

There were many challenges in evaluating the new methodology mainly that comparisons to the past were inevitable but not always rational given all the changes that occurred. Also, with the sheer volume of changes, it was difficult to isolate a single factor or two that explained changes in readership levels. The following table highlights many of the factors that could result in changes to audiences, with each of these factors ‘playing’ differently depending on the profile of each publication’s readers, both in terms of impact and magnitude.

Factors	Impact
Weighting	Differential impact by title depending on sample composition and extent of sample weights
Logos	Increases recall levels
Screening Changes	Potentially capture new demo graphic groups
Switch to Recent Reading	Increases readership levels compared to Issue Specific
Mobile Phone Only Households	Sample source may capture a different profile for this segment
18-24 Year Olds (River Sample)	Sample source may capture a different profile for this age cohort
65+	Sample source may capture a different profile for this age cohort
Oversample of High Income/Managers/Professionals	Over-representation may impact readership profile
Mode of Data Collection	Differential coverage, measurement or non-response error

Response Rate (Limited call-backs and no refusal conversions)	Increase variability in hard to reach segments of the population
Incentives (Readership Survey)	Potentially create differential bias
Offline Population	Under-represents a segment of the population who may be different from their online cohorts

As was the case with the pilot, a complete review of the national and individual market data was completed on the 6 month database and focussed on three key areas

- Study mechanics
- Sample of respondents
- Readership

Study mechanics

A full review of the study mechanics was undertaken. This involved assessing the recruitment steps, checking questionnaire logic, and reviewing weighting protocols. Once these initial steps were completed, a review of the tabulation process was conducted including a detailed review of the Average Issue Readership (AIR) calculation.

At the end of this due diligence process, no technical issues were identified. The mechanics followed closely to the protocols that were developed for the new study.

Sample

The sample was reviewed as a whole (both nationally and individual markets) and then more granularly among key demographic groups.

Sample consistency with the new method was an initial step to assess any potential variability in sample composition over time (that is the first quarter of January to March to the second quarter of April to June). The sample showed high consistency (within the 95% confidence interval) between the compositions of the Q1 and Q2 samples on 10+ variables (e.g., age, gender, occupation, education, income, household size, etc.).

Analysis was done to compare the national sample to known census data (Statistics Canada) on key demographics:

- Age: Sample under-represents 12 to 17 year olds versus population (which is not surprising for this demographic and is comparable to past research with this cohort) which is offset by an over-representation of 50+ years
- Gender: Higher proportion of males in the sample
- Education: Higher proportion of student in the sample (partly due to oversampling of 18 to 24 year olds) but sample is better educated (mainly at the university level) than the Canadian population
- Income: Personal and Household income reflect the population better as more lower income consumers are included in the sample (versus past methodologies)
    - It is noted that while income skews lower to the previous methodologies, the education profile matches previous data (i.e., sampling picking up educated people but later in their life where they are retired with lower income)
    - The matching of incomes of the sample to the population also indicates that the necessary oversample of higher income households was sufficient
- Home: Less home owners and more renters
- Language: While Canada has two official languages (English and French), there is a relatively large population of allophones (language spoken most is neither English nor French) and this methodology under-represents this group.

A cross-sectional analysis of selected key demographic groups was also done to help inform the interpretation of the new audience data relative to previous studies. These included…

1. *Mobile phone only households*: Still remains a targeted sample but screening now done via IVR and purposefully dialling mobile phone exchanges (as opposed to previous use of online access panels);
    - This groups skews younger, particularly in the 25-34 year range, with an income that tends to be lower. This age trend for these mobile only households matches benchmark data available on these households.
2. *18 to 24 year olds*: This cohort is now sampled differently with a targeted oversample using non-probability methods (and sample now delivers a representative proportion as opposed to previous under-representation)
    - Tend to have a lower household income which matches with census data.
3. *65+ years olds*: Although the sample delivers these consumers proportionately, the behaviour of this cohort in terms of media usage, public transit, or other activities is quite different from that seen in past research
    - Although this cohort is better educated than previous studies, their household income and personal income trend lower, which all correlate to a retired population. However, they also have higher internet use and public transit use than in the past (with the former obviously relating to the mode of the interview).

Given the differences in sample from previous audience research, it is not unexpected to see changes in media consumption behaviour in the new study. Differences by age were more pronounced - particularly for television viewing and internet usage - but these continue to build on the explanation of differences in audiences versus previous studies.

- Overall lower TV viewing (hours/week) across age, gender, and education
- Radio listening was generally comparable across various sub-groups
- Marginally higher Internet usage across the board but notably among the 65+ age group (although as mentioned already, this is a function of the data collection mode).

In summary, the sample coverage generally matches known census data from Statistics Canada although under-represents some of the new Canadians that have recently entered the country. The new methodology brings in people with a different demographic and behavioural profile compared to the previous studies, all leading to subsequent impacts on audience levels and profiles for different publications from the past.

Readership

As previously discussed, a large number of factors have changed from the past audience measurements of both newspapers and magazines and these will have an impact on estimates and the resulting reader profiles that limit any comparisons to existing data. While comparisons are inevitable, assessing the key drivers of the change is more useful in help aid discussions about the direction/magnitude of the change. The previous section focussed only on the sample composition but does begin to provide some direction as to what changes might be seen by publication; however, there are a number of factors related to the actual mechanics of the survey and data processing that would produce divergent results.

- Response rate
    - This is not a response rate study with rigourous controls on day of week dialling, callbacks, and refusal conversions as with the past research. Controls are placed on day of week recruitment but there is very little control on when a respondent will begin the self-complete survey. All of these will produce a lower response rate versus the past which does increase the potential for sample variability.
- Incentives
    - All respondents who complete a readership survey are incentivized (differential market incentives) and receive a further incentive for completing the product survey. This does compare to the past magazine readership survey but the newspaper survey did not incent respondents for their readership date (only product information). Using incentives can lead to differential bias, however, are necessary to maximize the completion rate (the pilot showed that more modest incentive levels hampered completion rates considerably).
- Screen-in changes
    - Magazines retain their initial P12M screen-in question at the brand level but the subsequent recency question now covers all four platforms (print, PDF, web, app) concurrently versus the previous print only approach. Digital only readers will no longer screen out.
    - Newspapers also retain their P3M screen in at the brand level but all four platforms are now asked concurrently versus sequentially.

- Both newspaper and magazines screen-ins included aided recall using the logo/masthead of each publication. Previous newspaper audience data did not benefit from the visuals as it was a telephone survey. Also, many of the past visuals included a single logo/title, but with the addition of digital properties some publishers have opted for a multi-logo visual.

- AIR Metric
  - Newspapers had been measured using Issue Specific but with the move to online data collection and lack of day of week controls, a switch to Recent Reading was logical.
  - The initial AIR model mirrored the same algorithm as magazines had been using which included those in who fall into one of two qualifying period for each magazine, with the second period being a mathematical factor approximating the likelihood of reading.
  - The model was updated for newspapers after subsequent review and that AIR people will only be based on the 'read yesterday' qualifying period and not include a factor of people who may have read it longer ago.
  - While this new AIR metric more closely aligns with that of the past metric, the relative differences in the context of the question would produce different results.
- Weighting
  - The previous newspaper and magazine audience measurements both included a multi-stage weighting scheme to align the results to known census data and represent the population. However, the same factors in each past study were not weighted. Specifically, personal income adjustments was imposed for the past magazine research but not newspapers. In the new study, the personal income adjustment would be included the complete sample.
  - Analysis was conducted both with and without the personal income adjustment and there were negligible differences.

Screen-in levels are susceptible to changes in interview procedures, sampling, questionnaire wording, and aided recall to name a few and all of these factors are in play, when comparing to past results. Not surprisingly, the average screen-in levels for each publication increased, 25% for newspapers and 40% for magazines. The unexpected element was that the incremental readership was coming from some of the demographic groups that had been under-represented in the past. In other words, readership in Canada has not increased but the methodology is capturing a different group of readers. Specifically these are…

- Younger consumers (<25 years) who are students/unemployed and have low household and personal income. This younger cohort was very under-represented in past studies and a more robust sample could provide different results.
- Older consumers (65+) who are retired and have a lower household/personal income. While these older consumers were well represented in the past, the previous methodologies would under-represent low income seniors (i.e., not living in houses) who appear to be behaviourally different.

As with screen-in levels, there were similar expectations of changes in both the level and composition of an Average Issue Reader. This would be the case because of the wider net being cast with the resulting sample, screen-in changes, and the Recent Reading change for newspapers. Overall increases in AIR for newspapers averaged 20% for magazines and 30% for newspapers and again much of the increase came from groups that were not well represented in the past, including…

- Younger (<35)
- Not employed full time: student (younger) or retired (older)
- Lower income (household and personal income)
- Rent home (versus own)

In summary, the vast changes in methodology and protocols should (and did) lead to different audience results across the board as we are now capturing some segments of the population that appear to be different/under-represented from the previous studies. However, not all changes can be quantified given the number of different factors that could be at play but many of the changes seem logical given the type of publication/reader profile/market.

Next steps

As a precursor for the release of the initial data to the market, various stakeholders were provided a top-line summary of the results through a number of meetings in early September 2015. These discussions centred on the differences (both in method/sample and audience data) as a path to move forward and accept the new data. And to this end, the industry will be releasing the first results of the new methodology (using a 6-month database) at the end of September. It is expected that more discussions will happen as stakeholders and the market delve deeper into the data.

Although the thorough review of the first 6 months of data collection did not yield any major issues, there were a number of initiatives that need to be evaluated and implemented for beginning of data collection in Q1'16.

1. *Measuring Digital Readership*: There are significant challenges in measuring digital readership based on recall as many consumers are not aware of how they exactly access digital content. In the survey pilot, potential confusion was definitely seen across platforms particularly between apps and digital replicas/PDFs. Suggested changes were made before the start of Q1'15 interviewing, which appears to have lessened some of the misattribution. However, there is still considerable confusion on digital platforms to the point that only a single NET digital measure will be initially released. Qualitative tests among consumers will be conducted in Q4'15 with the intention of using more consumer friendly language/visuals/aids to lessen the misattribution further. Additionally, there will be a review of how other countries manage digital readership, specifically with a focus on device calibration to 3rd party measurement.
2. *Data collection mode*: Given the under-coverage of some groups, there will be an investigation about the advantages/disadvantages for adding data collection modes. That is, what is the upside/downside to including both telephone and personal interviewing?
3. *Sample*: 12 to 17 year olds are under-represented in the sample and may over-exaggerate this cohort when weighting has been applied. Also, it is questionable that this group can accurately answer all questions, particularly those with weighting implications such as income. Further discussions among stakeholders will determine the value of this group in the sample moving forward.
4. *Response rate*: Will imposing a response rate target on the sample alter readership results or does a more rigourous dialling plan with greater callbacks/refusal conversions suffice?
5. *Language spoken*: Although we only measure English and French publications (and only offer the survey in those languages), many newer Canadians do not speak either of those languages at home but it is important to ensure we capture this growing demographic appropriately. Investigation of the merits of adding additional languages for the survey to be completed in or targeting these consumers in specific markets will be completed.
6. *Visual stimuli*: Does including colour logos (versus the current black/white) aid in recall? What about actual front covers of publications?
7. *Survey length*: The average completion time is 40 minutes for the readership survey and covers magazines and newspapers plus a myriad of demographics. The sheer length of the survey brings into question respondent fatigue and resulting data quality. This is to be an ongoing topic to ensure we only ask relevant questions.

The collection of print and digital audience data for Canada's newspaper and magazine industries had changed radically in the past year with a more fluid research methodology that delivers more data more frequently (and faster) to stakeholders than ever before. Although we are still in the initial stages of this journey, we must strive for continuous improvements to this audience measurement and the above initiatives put us on that right track.