MEASURING MAGAZINE READING VIA THE INTERNET: TESTING THE EFFECT OF NUMBER OF TITLES AND OTHER QUESTIONNAIRE DESIGN ISSUES

Jay Mattlin, Knowledge Networks
Bruce Goerlich, Starcom MediaVest Group

Modes of collecting readership data have evolved considerably since Politz's face-to-face TTB studies half a century ago. While magazine readership studies are still conducted in person with pencil and paper in some parts of the world, many readership studies are self-administered on paper through the mail, and others have taken advantage of new technology and have migrated to CAPI or to self-administration on a diskette. MRI's experiments with audio-CASI in the United States are familiar to those who have attended prior readership symposia as well (Baim, Arpin, and Appel, 1995; Baim, Frankel, and Arpin, 1997).

The advent of the Internet raises fresh opportunities for conducting readership research. As with mail surveys, surveys on the Internet have some of the advantages of self-administration, permitting respondents to answer at their own pace, perhaps with greater thoughtfulness and accuracy, and eliminating errors due to variance among interviewers. Data collected via the Web have some of the advantages of CAPI in that respondents and interviewers do not have to be relied upon to follow skip instructions. And Web surveys have one unique advantage over most other methods that have made them very attractive in the U.S: they can be done much more quickly than mail or face-to-face studies.

The main problem researchers have raised with using the Web to do survey research has been external validity or projectability. Web samples are usually considered inappropriate for general population surveys, because they are inherently biased towards those with Web access, especially frequent Web users and those who like to share their opinions on the Web. Moreover, there is no way to obtain a random sample of Internet homes or computers with Internet connections or e-mail addresses, because there is no sampling frame for any of these populations.

The founders of Knowledge Networks have attempted to rectify that problem. The company, a custom market research supplier which is barely two years old, has recruited a panel of over 25,000 households in the U.S. by phone from a RDD frame and sent them an appliance – a Web TV unit -- that gives them access to the Internet through their television sets. Knowledge Networks pays for the appliance and for panel households' monthly Internet access charges; the adult members of panel households, in return, are expected to complete one 10-15-minute questionnaire per week on their television sets through that Internet appliance. Thus, our samples are as projectible as RDD samples because they originated in the same way.

Nonetheless, collecting data through this new medium presents numerous challenges in attempting to measure readership of magazines reliably and accurately. How closely should the questionnaire hew to the country's standard magazine readership questionnaire? How many magazines can or should be included? In what order should they be presented? How many of them should be shown on a computer screen at one time? Should this new data collection tool be used to collect data that were impractical to collect through earlier methods? In short, to what extent does a new mode of collecting magazine data offer an invitation to revisit the structure of the data collection instrument, without compromising the quality or credibility of the data? These are the questions we addressed in the experiment we will be reporting on here.

GENESIS OF THE EXPERIMENT

This experiment grew out of Knowledge Networks' efforts to build a comprehensive database on the characteristics of our panelists to help us better serve our custom research clients. Knowledge about our panelists' television viewing, their use of the Internet, their health problems, the physical activities they engage in, their psychographic landscape, the cars they drive, and the magazines they read paint detailed portraits of them as individuals. Knowledge Networks develops questionnaires about each of these areas and sends them to panelists during weeks that they are not assigned a questionnaire for a client study. These data can be used by our custom research clients for selecting targeted samples, for enriching the data they are collecting from a survey, or for mining our database for insights about the consumer marketplace.

Our initial aim was to develop a magazine-reading questionnaire for our database that included questions on six-month reach, frequency of reading, newsstand buying, and subscriptions on as many magazines as possible. (Information about newspaper reading was already being collected in another survey.) However, we faced one key limitation: The interview could last no longer than 15 minutes. That's the maximum length we promise to panelists when they are recruited, and without our panelists' trust in us, our panel's existence would be jeopardized.

In order to maintain this length but still get information about 100 or so major magazines, we dropped the subscription and newsstand questions. Our final questionnaire consisted primarily of a six-month screening question, which was asked of all 106 titles, and a frequency-of-reading question, which was asked about all of the magazines that "screened in."

A word on terminology is important here: From this point onward, when we use the word "screen," we generally mean a part of a questionnaire that is displayed on a television screen. A "screen" in a Knowledge Networks questionnaire can consist of a single question, a grid with multiple items, a lengthy introduction to a question, or an image that is displayed as part of a question. If we are referring to the six-month reading question or six-month reading data, we will use the term "screen-in" or mention the phrase "six-month" explicitly.

Initial Questionnaire

In designing the initial magazine-reading questionnaire, we tried to follow the standard data collection instruments in the U.S. pretty closely. Our six-month screening question prompted respondents to think about all of the places that they may have read a magazine. Magazine logos were presented to respondents on a series of screens, and respondents were asked to indicate whether they had read each of the publications in the prior six months, had not read it, or were not sure. The frequency-of-reading question showed the logos of the magazines that had screened in and asked whether the respondent typically read 0, 1, 2, 3, or 4 of every 4 issues that are published.

Even though the basic structure of the six-month screening question was very similar to the industry standard in the U.S., our data collection platform necessitated a number of differences. An example of one of the screens from the questionnaire (Appendix 1) illustrates these differences:

1. First, we presented five to seven logos on each screen, rather than one logo at a time.

2. The respondent clicked on little circles in Yes/No/Not sure columns rather than putting logo cards in Yes/No/Not sure piles.

3. The logos were grouped in alphabetical blocks. Each screen showed the logos in one block. The order of the screens was randomized, as was the order of the magazines within each screen.

4. The logos were in color.

This initial foray into magazine measurement proved to be unsatisfactory to a number of internal and external constituencies:

1. Our potential magazine clients were generally more interested in identifying newsstand buyers or subscribers than frequent readers.

2. There was pressure on us to expand our original list of 106 titles.

3. Our analysis of the data from the first ten weeks of interviews in the fall of 2000 suggested that our respondents might be overclaiming their readership levels.

In order to assess the validity of our methodology, we compared our first 17,822 cases to MRI's audience data. MRI makes its AIR data available on MRI-Plus, a Web site, to those who register on the site. Since we had not collected recent reading data, we calculated AIR scores for the magazines in our survey from the frequency of reading question by applying the "conversion rates" for frequency of reading within publication frequency derived by David Napior and Cheryl Brink in their paper for the 1997 Symposium (Napior and Brink, 1997). These conversion rates represent the probabilities of a respondent's reading a recent issue of a magazine, based on the number of issues out of four that a respondent claims to read and the magazine's publication frequency.

We were encouraged to find that among the 89 magazines in our study that were included in MRI's survey, the correlation between our derived audience estimates and MRI's published recent-reading-based audience estimates was .97. This strong linear relationship between these two sources of data indicated to us that our methodology had captured the true underlying variability across magazines in their audiences.

However, we also found that our derived AIR numbers were an average of 81% higher than MRI's magazine audience numbers. This pattern is similar to the results of Nielsen's HOME*SCAN experiment, reported at the 1993 Symposium – a high correlation across magazines, but with higher levels of reading across the board (Joyce and Coffey, 1993).

A number of possible explanations were advanced to explain the discrepancy:

- Our sample: Our sample is generated through RDD, whereas MRI's is an area probability sample. However, our sample comes close to mirroring the U.S. population in most respects and is weighted to compensate for any differences between the sample's demographic characteristics and those of the U.S. population.

- Comparability of measures: Perhaps the comparison between the two sets of audience estimates was inherently problematic, because we were comparing an estimate based on frequency of reading to an estimate based on recent reading. However, this difference alone cannot explain the higher levels of issue reading in our study. Baim, Frankel, and Arpin reported in their 1997 paper on the audio-CASI experiment that an average of 15 magazines had "screened in" in their control sample (Baim, Frankel, and Arpin, 1997). In our sample, the average number of magazines that respondents reported reading in the prior 6 months was exactly the same, even though we were asking about fewer than half the number of magazines in MRI's survey. While it is true that the 106 magazines we picked for our study were generally the higher-circulation magazines, the screen-in rates in our study were inevitably higher than MRI's.

- Number of titles: It has been accepted as conventional wisdom that asking about larger numbers of magazines yields lower reported readership levels for individual titles than asking about smaller numbers of magazines. Perhaps a larger number of magazines intimidate respondents into being more cautious about the number of magazines they report that they had read, or conversely, perhaps a smaller number of magazines encourages respondents to screen in titles in order to appear literate. MRI respondents can see how many magazines they are being asked about, because they can see how big the logo deck is. In Knowledge Networks' questionnaire, the number of magazines wasn't mentioned; in the absence of this information, respondents might have assumed initially that there were even fewer than there actually were.

- Number of titles per screen: Any effect due to the lower number of titles overall may have been exacerbated by the small numbers of titles that appeared on each screen (only five to seven per screen).

- Data collection mode itself: Perhaps interviewing respondents on the Web inevitably produces readership levels that exceed norms from interviewer-administered face-to-face interviews simply due to the mode itself. Ours is certainly not the first study to report an increase in reading levels when a more electronically advanced data collection method is employed. MRI's Audio-CASI experiment, Nielsen's Home*SCAN test, the British NRS' transition to CAPI, even Intelliquest's pilot of a disc-based questionnaire all found higher reading levels than the prevailing control studies to which they were compared (Beeson, 1993; Pinnell and Arpin, 1995).

RESEARCH QUESTIONS

Our effort to evaluate these explanations, to redress some of the limitations of the initial questionnaire, and to attempt to take fuller advantage of our data collection medium led to the experiment undertaken by Knowledge Networks and the Starcom MediaVest Group this past spring. We had hoped to find the optimal way, using our platform, to expand the list of magazines to 200 and include information on newsstand buying and subscriptions, while improving the accuracy of our data. This experiment was intended to address the following research questions:

- Whether changing the format of the six-month screening question from a series of yes/no/not sure response options to "check-all-that-apply," as well as mentioning the number of logos to be shown, reduces over-reporting of magazine reading (**Issue #1**)

- Whether, and the extent to which, greater numbers of titles reduce reported readership levels (**Issue #2**)

- Whether self-reported readership of individual magazines can be improved by preceding the six-month screening question with a context-setting question about the number of magazines the respondent reads (**Issue #3**)

- Whether it is possible to produce credible estimates of magazine audiences by asking respondents to name the magazines they have recently read, unaided. (**Issue #4**)

- Whether selectively asking respondents about only the categories of magazines they report reading reduces over-reporting of magazine readership. (**Issue #5a**)

- Whether presenting the magazine logos in clusters within the same genre, as is done in the British NRS, improves reporting of readership. (**Issue #5b**)

- Finally, whether calibrated frequency of reading, recent reading, or the First-Read Yesterday (FRY) technique produces the most credible audience estimates for magazines through this data collection mode. (**Issue #6**)

This experiment appeared to us to be a good opportunity to explore the fundamental issue about which question sequence is the most appropriate for producing magazine audience estimates when collecting readership data through the Internet. A number of

print audience research services (e.g., Monroe Mendelson in the U.S.) base their print audience estimates on frequency of reading, a question we were interested in including in this survey, as we did the initial one, for possible custom research applications. The recalibrated frequency-of-reading methodology also has the virtue of being well-suited to a self-administered survey that is completed at a time of the respondent's choosing, because the reference period is general (Lindberg, 1997). This was considered the "control" AIR estimation algorithm here. Recent reading is the measure most directly comparable to the leading source of print audience data in the U.S. And FRY, while used relatively rarely, possesses two oft-recognized virtues:

(1)   It greatly reduces the threats of parallel and replicated reading;
(2)   It is the least taxing of respondents' memories.

DESIGN OF THE EXPERIMENT

Five different questionnaires were developed to investigate these research questions. These five questionnaires covered four different treatments, as explained below. All five questionnaires had the following comment elements, intended to investigate Issues 1 and 6:

- **A six-month screening question** that was very different from the one used in the initial test. This time, the introduction to this question informed the respondent about the number of magazines he or she would be asked about and the number of screens of logos he or she would be shown. Perhaps more important, the Yes/No/Not sure format was eliminated. An average of only 1.2% had answered "not sure" for the 106 magazines in the initial survey, and so this option was thought to have little practical value.

  This was replaced by a "check all that apply format" in which respondents were shown logos in a series of groups and asked to check the ones in each group that they had read in the prior six months (See Appendix 2 for example). An advantage of this format, we believed, is that it allowed us to fit a far greater number of magazine logos on a single screen. There are two benefits to doing this: First, more logos per screen, we hypothesized, would reduce over-reporting. And second, more logos per screen meant fewer screens were needed to ask the six-month reading question for a given number of magazines, which allowed room for a greater number of additional questions.

- **A frequency-of-reading question**. A change was made here, too. In the initial study, this question was asked about all of the screened-in magazines at once, in a single screen. In this test, the magazines were divided into groups by publication frequency, and the frequency-of-reading question was administered to each group separately.

- **A recent reading question**. There were two forms of this question, one for weeklies, and one for non-weeklies. The form for weeklies was as follows: "When was the last time you read or looked into an issue of each of these magazines?" The answer choices were:
  - Earlier today
  - Yesterday
  - In the last 7 days
  - More than 7 days ago

  The question about non-weeklies was the same, but the answer categories were expanded to . . .
  - Earlier today
  - Yesterday
  - In the last 7 days
  - In the last 30 days
  - More than 30 days ago

  The analyses of recent reading data reported below include data on weeklies and monthlies only.

- **A First-Read-Yesterday (FRY) question**: In order to test the viability of FRY on our platform, we included the following question that was asked of all those who checked "yesterday" in the recent reading question: "Think about the magazines you read yesterday. Was yesterday the first time you read that specific issue of the magazine?"

This experiment sought to measure six-month reading and average issue reading for 200 magazines altogether. Since all four of the treatments contained all three of the possible magazine audience estimation questions, the elements that varied across the four treatments were:

- The numbers of magazines included in the questionnaire;
- The way in which the magazines were screened in;
- Whether the questionnaire included questions about the magazines to which the respondent subscribed, those they had bought on the newsstand in the prior six months, and the amount of time they had spent with the issues of magazines they had read in the prior week.

The four different treatments are described below:

- **TREATMENT A: The "Control" versions**: There were two control questionnaires, each covering a randomly-selected subset of 100 magazines out of the 200 we sought to measure. This was considered the control condition, because 100 was the number of magazines included in the initial study we had conducted. Aside from the list of magazines in each questionnaire, the two control versions were otherwise the same. In each of the two control questionnaires, there were five screens for the six-month reading question, each containing 20 magazine logos and a box for "none of the above," laid out in three columns of seven (see Appendix 2). These two versions, when combined, represent one experimental condition in the four-cell test.

- **TREATMENT B: The "Full-Deck" Version**: This version contained all of the 200 selected titles. The format and structure of the six-month screening question was otherwise the same as in the control versions. The six-month reading question for this version consisted of ten screens, each of which contained 20 magazine logos. Besides number of titles, the only other differences between this version and the control versions were that this version did not include the subscription, newsstand, and "time spent" questions. These questions were omitted in order to keep the full-deck version to a tolerable length for our panelists. This version allowed us to explore issue #2.

- **TREATMENT C: The "80-Title Open-End" Version**: The third version of the questionnaire, designed to address Issues 2, 3 and 4, contained three experimental elements:

    1. First, respondents were shown only 80 logos in the six-month screening question, a random sample of the 100 logos in one of the two control versions (Issue #2);

    2. Second, the six-month screening question was preceded by a general question about the number of magazines the respondent reads. This question was partly intended to frame the subsequent six-month screening battery so that respondents would be subtly prompted to make their answers to the six-month screening question consistent with their answer about the number of magazines they read. It was thought that by forcing respondents to think first about the volume of their magazine reading, they would be less likely to overclaim readership of individual magazines (Issue #3). This question served another purpose as well (noted below).

    3. Finally, we attempted something new to take advantage of the capabilities of our platform: After asking about the roster of 80 magazines, we reminded respondents of the number of magazines they had reported that they generally read and suggested that our list of 80 might not have included all of the magazine they had read in the prior 6 months. Respondents were then asked to record the names of all of those magazines we missed – the "other" magazines they had read in the prior six months. Respondents were given the opportunity to type in up to ten magazines, using their Web TV keyboards.

    This was an experiment to determine whether it was possible to obtain audience estimates for small-circulation titles by asking respondents to name the less-known magazines for which they might have a strong affinity (Issue #4). Though we realized that this approach might produce fewer screen-ins for the unlisted magazines, the underlying hypothesis was that the magazine titles volunteered by respondents would be more likely to have been read frequently and recently. Lower screen-in rates would be balanced by higher read-to-screen ratios so that AIR estimates, it was thought, would be robust.

In Versions A, B, and C, the logos in the six-month screening question were presented in one of two orders – alphabetical order or reverse-alphabetical -- to which respondents were randomly assigned.[1]

- **TREATMENT D: The "Category-Filter" Version**: This version was aimed at Issues 5a and 5b. It was designed to test a strategy for obtaining more accurate information about 200 magazines in a single questionnaire without inducing egregious levels of respondent fatigue and annoyance by filtering out genres of magazines that respondents never touched. The idea behind this test was that there was no purpose served in asking certain people whether they had read 15 different automobile magazines in the past six months when a single question about whether they had read any

---

[1] For technical reasons, there was a delay in the implementation of this rotation, resulting in a greater number of respondents being assigned the alpha order rotations. As noted later, the data were weighted to redress this imbalance.

automotive magazines during that time period would spare these respondents the effort of perusing 15 unfamiliar logos of automobile magazines. Moreover, it was believed that withholding the logos of magazines respondents were highly unlikely to have read would reduce accidental or haphazard screen-ins of attractive or familiar logos. In addition, it was thought that grouping the magazines by genre for the six-month screening question, as is done in the EML Grouped Titles Method in Britain, would improve the accuracy of respondents' self-reports by enabling them to focus more closely on one group of similar titles at a time. It was also hypothesized that grouping titles within genres would help to reduce title confusion, as has been found in the British NRS, because respondents would be able to see the logos of magazines with similar titles side-by-side (Meier and Finch, 1993).

In this version, the six-month screening question was preceded by a series of items asking respondents whether they had read any magazines in specific categories in the preceding six months. Respondents were subsequently asked the six-month screening question only about magazines whose *categories* they reported having read in the prior six months. Each screen of the six-month reading series showed the magazine logos within a particular category, either alphabetically, or reverse-alphabetically (in a random rotation). These screens were shown to the respondent in random order. They tended to have fewer than the 20 logos per screen in the other three versions. Some of these screens – the ones for newsmagazines and travel magazines – contained just three or four logos.

We developed 20 of these magazine categories (See Appendix 3). Each magazine was classified in only one category in order to equalize opportunities to be asked about each magazine. However, there were 19 magazines for which we were unable to define a category that we felt respondents would understand ("regional" magazines such as *Texas Monthly* and *Midwest Living*; "general interest" magazines such as *The New Yorker*) or whose category definitions we were concerned might offend respondents ("mature market" magazines or "tabloids"). All respondents were asked the six-month readership question about these 19 magazines.

This methodology reflects recent trends in media consumption and media planning, with increasing fragmentation of audiences for all media and media vehicles. Ad agencies frequently evaluate magazines in "clusters" with similar reader characteristics (Singer 1997). The ability of individual magazines to reach specific target markets is often viewed as more relevant for media planning purposes than the size of their readership bases. In recent years, questions have even been raised in the U.S. about the value of total audience measurement (Adelman, 1997). This experiment represents a small step in the direction of focusing magazine research on the measurement of more narrow audience targets.

The ways in which the four test versions of the questionnaire differed are summarized in the table below:

	Version A: "Control Versions"	Version B: Full-Deck Version	Version C: 80-Title Open-End Version	Version D: Category-Filtered Version
Number of logos included in 6-month screening question	100 each	200	80	200 max., though numbers shown to individual respondents varied
Questions prior to screening question	4 introductory questions[2]	4 introductory questions[2]	4 intro Qs, plus Q. about total number of magazines read	4 intro Q's, plus Q. about readership of 20 types of mags in prior 6 months
Open-end Q. about other magazines read in prior 6 mos.	Not asked	Not asked	Asked, with space for up to 10 responses	Not asked
Questions about subscriptions and newsstand buying	Asked of all screened-in magazines	Not asked	Asked of all screened-in magazines	Not asked
Recent reading and read yesterday[3]	Asked of all screened-in magazines	Asked of all screened-in magazines	Asked of all screened-in magazines	Asked of all screened-in magazines
Time spent with recently-read issues	Asked of all mags read in last 7 days	Asked of all mags read in last 7 days	Not asked	Not asked

Administration of the Study

The five versions of the survey were sent to a total of 6,700 panelists between May 25 and June 1, 2001. Altogether, 5,443 of

---

[2] The first of these questions was intended to be a gentle introduction to the topic – what the respondent would do with an additional hour a day. This was followed by questions about the number of magazines to which the respondent subscribes, the frequency with which she buys magazines at the newsstand, and the frequency with which she reads magazines.
[3] Since "last two weeks," "last three weeks," and "last two months" were not included as response categories in the recent reading question, the data for bi-weeklies, tri-weeklies, and bi-monthlies were excluded from all recent reading analyses.

them responded (81%) in the two weeks before we closed out the survey (although some of the assigned panelists responded afterward). The sample sizes for each version are shown below:

Version A: Control (both sub-versions)   1,661
                    Sub-version 1   818
                    Sub-version 2   843
Version B: Full-Deck Test   1,116
Version C: 80-title open-end Test   1,327
Version D: Category-specific test   1,339

The survey data were weighted to account for differentials in probabilities of initial selection of the panelists at the RDD recruitment phase, for demographic differences between the sample and the U.S. population as a whole, and for the greater numbers of respondents who completed the alphabetical-order version than the reverse-alphabetical version. The order in which the logos were shown did make a difference in the magazines' screen-in rates: For Versions A through C, the first 40 magazines in alphabetical order and the last 40 in alphabetical order were, in general, significantly more likely to screen in when they appeared in the beginning – especially in the Full-Deck Version.

	Difference between percent read in the last six months when magazines appeared in the first two screens from when they appeared in the last two screens			
	Version A1	Version A2	Version B	Version C
First 40 Magazines in alphabetical order:				
Mean difference	+1.3%	+0.5%	+1.3%	+0.4%
Standard Error	0.3	0.4	0.3	0.2
Ratio	4.4*	1.3	4.3*	1.7
Last 40 Magazines in alphabetical order				
Mean difference	+0.4%	+0.8%	+1.4%	+1.0%
Standard Error	0.3	0.4	0.4	0.4
Ratio	1.2	2.1*	4.0*	2.7*

*p < .05

RESULTS

Each of the research questions raised earlier will be discussed in turn. A summary table comparing the six-month screening rates and various methods for deriving AIR across the four versions is shown in Appendix 4.

Issue #1: Switch to Check-All That Apply

As we had expected, shifting the six-month reading screens from a yes/no/not sure format for every magazine to a check-all-that-apply format for groups of magazines yielded fewer screen-ins and lower average AIR scores. For the 99 magazines that appeared in both the initial version that preceded the experiment and the two control versions of the experiment, the average number of titles respondents reported reading in the prior six months fell from 14.4 to 8.8, a 23% decline.[4] The average of those 99 magazines' AIR scores, calculated with the recalibrated frequency-of-reading scale, dropped 48% (from an average of 6.39% to 3.29%).   By not forcing respondents to provide an answer for every magazine and by showing them five screens of logos instead of 18, we made this portion of the interview easier for respondents, perhaps reducing any pressure they might have felt to maximize the number of magazines that they reported reading in the prior six months. In addition, letting them know in advance the number of logos they would be shown might have reduced their incentive to report readership of lots of titles for the sake of appearing to be well-read. (The way in which the test was structured made it impossible to separate the effect of check-all-that-apply from the effect of notification about the number of magazines included in the survey.)

---

[4]Of the 99 magazines in the original questionnaire included in this test, 48 appeared in one of the two control versions, and 51 appeared in the other control version.

Issues #2 and #3: Number of Magazines Included

The number of magazines included in the six-month screening question did, as we had anticipated, affect the proportion of magazines that screened in. The greater the number of magazines that were asked about, the smaller the percentage of those magazines that respondents reported that they had read in the prior 6 months:

Version	Number of Magazines Shown in six-month Screening Questions	Average number of magazines screened in	Projected number of screen-ins if all 200 magazines had been listed in the same questionnaire*	Index vs. control versions
B (Full-Deck)	200	11.5	11.5	95
A (Control)	100 (Average of the two control versions)	6.0	12.1	100
C (80-title & open-end)	80 – Adjusted to account for greater-than-average readership of these 80 magazines in the other 3 versions[5]	5.3	13.3	111
C (80-title & open-end)	80 – Unadjusted	6.1	15.2	126

*Average in the column at left multiplied by 200 and divided by the number of magazines covered in the version

The relationship between number of magazines listed in the questionnaire and the average number screened in is curvilinear, rather than linear. The 11% gap between 80 listed magazines and 100 is double the 5% gap between 100 magazines and 200 in the average number of screen-ins per listed magazine. It therefore appears that the number of magazines included in our original study can account by itself for only a very small portion of the elevated reading levels we found for those 106 magazines, in comparison to MRI.

It also appears that the question in the 80-magazine version about the number of magazines you read, which preceded the six-month screening question, did not reduce screen-ins for the 80 magazines that listed in the questionnaire. While it is impossible to tell whether it may have actually served to boost the number of screen-ins in the 80-magazine version, it is unlikely that it tempered the impact of the smaller number of listed magazines.

The number of magazines listed in the questionnaire had little impact on read-to-screen ratios, based on frequency of reading. Among the 80 titles listed in all three versions, read-to-screen ratios ranged from .38 to .40, and there was no linear relationship between number of magazines listed and read-to-screen ratios. As a result, the distance between the version with 80 listed magazines and the two "control" versions with 100 each persisted with respect to AIR, based on the recalibrated frequency scales. For the 80 magazines listed in all four of the "treatment" and "control" conditions, AIRs were an average of 14% higher in the 80-magazine version than in the "control" version. The difference between 100 and 200 magazines diminishes at the "average issue" level, however, to less than a tenth of a percentage point.

	Average of 200 Magazines		Average of 80 Magazines covered in all four treatments		
	Version B: Full-Deck	Version A: Combination of two control versions	Version B: Full-Deck	Version A: Combination of two control versions	Vers. C: 80-title version
Number of magazines covered in questionnaire	200	100	200	100	80
Percent who read each magazine in the past six months	5.74	6.03	6.51	6.98	7.62
Read-to-screen ratio, based on recalibrated frequency scale	.39	.39	.40	.38	.39
Estimated AIR percentage based on recalibrated frequency scale	2.22	2.28	2.55	2.58	2.94

So, the number of magazines included in an audience questionnaire on our platform does appear to affect estimated screen-in rates and, concomitantly, audience levels, but above 100 magazines, the effect is relatively slight.

---

[5] The adjustment was derived by (a) calculating the average percentage of screen-ins accounted for by those 80 magazines in the other three versions (0.457), (b) dividing the unadjusted average number of screened-in magazines (6.1) by the average percentage calculated in (a), and then (c) multiplying that ratio by 0.4 -- the proportion of screen-ins that those 80 magazines should have accounted for out of 200 magazines (80/200) if the screen-in rates of those 80 magazines had mirrored the average screen-in rate for all 200 magazines.

Issues #3 and #4: Open-ended responses

The results of the open-end experiment in Version C are a testament to the multiplicity and variety of magazines that are published in the United States today. Altogether, over 1,100 different titles were written into the blank spaces at the bottom of the screen by the 1,327 people who answered that version of the survey. Sixty-two percent of the respondents filled in one or more magazines, with an average of 3.2 filled-in magazines per person among those who typed in at least one magazine.[6]

And yet, our experiment revealed that a solicitation of the names of magazines read in the last six months in an open-ended question is no substitute for the traditional closed-ended queries with logo prompts, even in our new, more interactive data collection platform and even if framed by questions intended to spur respondents to think of all of the magazines they might have read. As expected, the 120 magazines not listed in Version C screened in at a much lower rate than in the other versions. And, as expected, the read-to-screen ratios for the unlisted magazines were much higher than they were in Versions A and B, where they were listed. However, the higher read-to-screen ratios were not sufficiently high to compensate for the much lower screen-in rates. Estimated AIRs for all 200 magazines based on the open-end methodology are therefore far lower than they were for the two versions in which all 200 magazines were listed.

If all of the magazines written in by respondents are added to those that they checked that they had read in the past six months in Version C, the average number of magazines read per respondent during the screen-in period would rise to 8.1. This is 30% less than the average of 11.5 magazines read in the past six months by those who answered the full-deck version, even though the 8.1 includes over 900 titles written in by respondents that were not even covered in the full-deck version. If those 900 titles are excluded, the average number of screened-in magazines among the 200 covered in this experiment falls to 7.0. So, if we had used the open-end responses and the closed-ended responses in the 80-title version to estimate six-month reading levels for all 200 magazines, the average screen-in percentage for those 200 magazines would be 39% less than the average for those 200 magazines in the 200-title treatment.

The average read-to-screen ratio for the combined set of closed-end and open-ends was 41% higher than the average read-to-screen ratio for the closed-ended versions. However, these higher read-to-screen ratios were not able to overcome the version's lower screen-in rates. The average estimated AIR, based on frequency of reading, was 37% lower for the combination of open- and closed-ends than for the control version.

	All 200 Magazines, including write-ins in the 80-logo version		
	Average Percent Read in the Past Six Months	Average Read-to-Screen Ratio (based on frequency of reading)	Average Estimated AIR, based on frequency of reading
Version A: Control versions combined	6.0	.39	2.3
Version B: Full-deck	5.7	.39	2.2
Version C: 80 titles, plus open-ended responses to six-month reading question	3.5	.55	1.4

An analysis of the 120 magazines not explicitly listed in the 80-title version (see table at the top of the next page) reveals why the higher read-to-screen ratios for this version could not compensate for the version's screen-in deficit. The average percentage of respondents who wrote in that they had read each of these 120 magazines in the prior 6 months was 86% lower than the average six-month screening percentage for these 120 magazines in the full-deck version. For these 120 unlisted magazines, reaching the AIRs of the control version would have required read-to-screen ratios that exceeded 2.0, which is neither theoretically nor practically possible.

So, even though the read-to-screen ratios for these 120 magazines in Version C were an average of 70% higher than in Versions A and B, the average AIR for these open-end-mentioned magazines did not come close to their average in the control version.

	120 Magazines not explicitly listed in the 80-logo version		
	Average Percent Read in the Past Six Months	Average Read-to-Screen Ratio (based on frequency of reading)	Average Estimated AIR, based on frequency of reading
Version A: Control versions combined	5.4	.39	2.1
Version B: Full-deck	5.2	.39	2.0
Version C: 80 titles, plus open-ended responses to six-month reading question	0.7	.66	0.4

The under-reporting of reading in the open-end condition, in comparison to the closed-end condition, persisted for every publication frequency.

---

[6] Some respondents filled in the names of magazines that were already among the 80 listed in Version C of the questionnaire. This comparisons in this section include these additional mentions of listed magazines.

It should be noted that it would not have made a difference if we had allowed respondents to enter more than 10 "other" magazines. Two percent of the sample filled in all ten of the open magazine slots. Even if all of these heavy magazine readers had gone on to write in another 10 magazines, the average number of six-month screen-ins would have increased by 0.2 magazines, hardly enough to make up for the deficits shown above.

While the open-ended question may serve to elicit the names of new magazines to include in the set of logos to show in subsequent waves and may also serve as a helpful measure of reader attachment to publications, they appear to fall short in the service of quantitative audience estimation.

Issues #5a and #5b: Filtering and Grouping by Categories

The experiment to streamline the survey by preceding the six-month reading question for individual magazines with a question about readership of categories of magazines in the prior six months appears to have been more successful than the open-end experiment in achieving its purpose. The filter questions cut down substantially on the number of magazine logos respondents were asked about and the number of screens filled with magazines logos that they were obliged to examine. Even though this version of the questionnaire included all 200 of the titles listed in the Full-Deck version, respondents to the category-filtered version of the questionnaire (Version D) were asked the six-month readership question about an average of just 75 magazines on 8.5 screens. (These averages include the two screens of logos of the 19 titles that were shown to all respondents.)

It is not surprising, given the smaller numbers of logos to which these respondents were exposed, that the average number of magazines these respondents reported reading in the prior six months was lower than the average numbers reported for the traditional, unfiltered versions with the same numbers of titles. However, given one possible concern about this methodology -- that it would eliminate scores of titles from a respondent's further consideration because the way in which we classified magazines did not fit the respondents' sense of how the magazines should be classified – the fall-off in screen-ins was relatively moderate. Even though respondents to the category-filtered version were shown 60% fewer logos than respondents to the Full-Deck version, they reported reading just 7% fewer magazines in the prior six months. They reported reading 11% fewer magazines in the prior six months than the respondents to the 100-title "control" versions.

Version	Treatment condition	Average number of magazines screened in	Index. Vs. Control
A: Control	Combination of two versions with 100 magazines each	12.1	100
B: Full-Deck	200 titles, no category filter or grouping	11.5	95
C: Category-Filtered	200 titles with category filters and groups	10.8	89

The category-filtering methodology also produced lower read-to-screen ratios (derived through the recalibrated frequency-of-reading scale), thereby driving estimated AIRs even further below the levels for the control versions.

	Average Percent Read in the Past Six Months	Average Read-to-Screen Ratio (based on frequency of reading)	Average Estimated AIR, based on frequency of reading	Index of Estimated AIR vs. Control*
Version A: Control versions combined	6.0	.39	2.3	100
Version B: Full-deck	5.7	.39	2.2	97
Version D: 200 titles, with category filters and groups	5.4	.35	1.9	85

* Ratio of the averages in the prior column to 2.3

The category-filtering methodology did not appear to give weeklies an advantage over monthlies and bi-monthlies, with the conspicuous exceptions of *Parade* and *USA Weekend*, two Sunday newspapers supplements. The respondents who completed the category-filtered questionnaire reported lower levels of readership than respondents to the non-category-filtered versions for all publication frequencies, once these two Sunday supplements have been removed from the comparison. This pattern held true for both average issue readership and for six-month reading (shown in the table below).

	Average Percent Read in Last 6 Months			
	Control Versions	Full- Deck Version	Category-Filtered version	Index vs. Control*
Weeklies (not including *Parade & USA Weekend*)	11.7	11.3	10.7	92
Bi-weeklies	4.7	4.8	4.6	98
Tri-weeklies	11.8	10.5	8.5	72
Monthlies	5.7	5.4	4.9	86
Bi-monthlies	2.5	2.4	2.2	88

* Ratio of the averages in the Category-Filtered column to the Control Versions column

Another concern we had had about this methodology was that the 19 uncategorized magazines would receive favored treatment because their screen-in rates and AIR figures would be based on all respondents, not just those who acknowledged that they had read magazines in a particular category. This concern does not appear to have been warranted. The 19 uncategorized magazines enjoyed higher reading levels, both in the prior 6 months and for an average issue, than the other 179 magazines (i.e., not including *Parade* and *USA Weekend*) in every version of the questionnaire (see table at the top of the next page).

Though the sizes of the gaps between those 19 magazines and the other 179 were slightly wider in the category-filtered than in the control versions, the differences between the two versions were not at all close to being statistically significant with respect to six-month reading and recent reading. The gaps were about the same in the category-filtered version as in the Full-Deck version.

	Control Versions	Full-Deck Version	Category-filtered version
Average percent read in past 6 months			
19 Individually-Listed Magazines			
Magazines in Categories*	7.1	7.1	6.6
Difference	5.9	5.6	5.1
	+1.2	+1.5	+1.5
Estimated AIR percent, based on frequency-of-reading			
19 Individually-Listed Magazines	3.0	3.0	2.9
Magazines in Categories*	2.2	2.1	1.8
Difference	+0.8	+0.9	+1.1
Average Recent reading percent (weeklies and monthlies only)			
19 Individually-Listed Magazines	4.9	5.0	4.6
Magazines in Categories*	3.5	3.3	3.0
Difference	+1.4	+1.7	+1.6

*Note: *Parade* and *USA Weekend* excluded

Among the 20 magazine categories, this methodology did appear to favor some of them and to put others at a relative disadvantage. In comparison to the control and Full-Deck versions, six-month screening rates and average-issue reading rates tended to be higher in the category-filtered version for three types of magazines – food and cooking, in-flight, and Sunday newspaper supplements. On the other side of the ledger, seven magazine genres elicited consistently at least 15% lower six-month screening and average issue reading levels in the category-filtered version than in either of the two versions that showed the full set of magazines. These differences held whether AIR was calculated through recalibrated frequency scales or through recent reading (See Appendix 5):

- Hunting and fishing (65% lower frequency-based AIR than the average of the other two versions)
- Women's interests (39% lower)

- TV, music, or entertainment (38% lower)
- People and personalities (35% lower)
- Men's interests (31% lower)
- Science and nature (25% lower)
- Sports (24% lower)

Four of these categories – women's interests, TV/Music/entertainment, people and personalities, and men's interests -- are relatively broad; this might be the reason that they suffer from relatively low readership levels in the category-filtered methodology. Seeing one of the category names might not trigger a memory of a specific magazine in the same way that seeing a particular logo would. For the other three categories with readership deficits, the relatively low averages might be attributable to respondents' perceptions that some of the individual magazines we placed in these categories don't belong in those categories.

Herein lie two notes of caution about the category-filtered methodology: First, it appears that the estimates it generates are relatively close to those generated by more traditional direct questioning for more narrow and well-recognized categories, but broader categories should be avoided. Second, magazines whose designations are not clear-cut would be better left in the list of uncategorized magazines whose logos should be shown to all respondents (which, this experiment indicates, does not appear to give those magazines, on average, any consistent advantage over the ones that remain in categories.) Overall, our experiment suggested that the adoption of this technique should be preceded by extensive developmental research among readers of publications in a number of categories to ensure that the category names bring to mind the magazines that they are intended to represent.

Issue #6: Questions on which to base AIR

Perhaps the two most striking conclusions we drew from this experiment were that, of the three algorithms we tested for estimating magazine audiences with the Web TV platform, (1) recent reading produced the highest estimates and (2) FRY produced the lowest.

	AVERAGE READ-TO-SCREEN RATIOS			
	Control Version	Full Deck Version	Category Filter Version	80-title/ Open-end version: Listed logos only[1]
Recalibrated Frequency of Reading	.39	.39	.35	.39
Recent Reading[@]	.55	.56	.55	.59
First Read Yesterday	.32	.35	.23	.39
	AVERAGE AIR ESTIMATES (in percentages)			
Recalibrated Frequency of Reading	2.3	2.2	1.9	2.9
Recent Reading[@]	3.6	3.5	3.1	4.9
First Read Yesterday	1.9	1.9	1.5	2.0

Number of magazines . . .
 	
  For Recalibrated Frequency
   and FRY Estimates
  For Recent Reading Estimates
	
	198
	156
		
		198
		156
			
			198
			156
				
				78
				61

@ Weeklies and monthlies only
Note: Parade and USA Weekend excluded
[1] See Table A in Appendix 4 for read-to-screen ratios and AIR estimates for the combination of 80 listed logos and 120 open-end mentions in Version C. See Table B in Appendix 4 for comparison of the four versions based only on the 80 magazines listed explicitly in the 80-title version.

The FRY estimates were simply too low to be credible. Less than 15% of the respondents to each questionnaire version answered that they had read *any* magazines for the first time yesterday. The FRY estimates were also far less stable than the others. For every one of our versions, close to half (49%) or more of the FRY estimates were 0. At the same time, a small number of the magazines in every version ended up with unadjusted read-to-screen ratios, based on FRY, which exceeded 1.0.

Two possible explanations for the failure of FRY occurred to us:

- Because our interviews are self-administered, we generally have no control over the day on which the respondent decides to complete the survey. Most of our surveys are completed on weekends. Therefore, FRY projections would over- represent magazine reading on Friday and Saturday. The validity of FRY estimates depends on a uniform distribution of measurements across the days of the week. (It might not have helped us to weight the data to smooth out the distribution of days of the week that the survey was answered, because there might be a correlation between the days that respondents filled out the surveys and the days of the week that they read magazines.)

- Respondents may have had trouble understanding our FRY question. If other researchers are able to craft a FRY question that is clearer to respondents in a Web-administered survey, then they might have more success than we did in generating realistic FRY estimates.

Appealing as this methodology was to us from a theoretical perspective, we can understand why so few countries have adopted it.

Read-to-screen ratios for recent reading were substantially higher than the read-to-screen ratios for recalibrated frequency of reading across all versions. The four versions' recent-reading-based read-to-screen ratios were fairly close, all within 10% of each other. The category-filtered version, whose frequency-of-reading-based read-to-screen ratios were lower than those of the other versions, had virtually the same recent-reading-based read-to-screen ratios as the other versions. Nonetheless, because of their lower six-month reading rates, respondents to the category-filtered version ended up with lower levels of recent reading than respondents to the other versions. The 80-title version, whose six-month reading rates were the highest among the four versions, also had the highest read-to-screen ratios, based on recent reading. This remains true when the comparison to the other versions is based only on the 80 magazines that were listed in Version C (See Table B in Appendix 4). It therefore appears that showing fewer than 100 titles may lead to higher reported levels of recent reading as well as higher six-month reading rates.

Comparisons of the audience data produced by our three estimating algorithms to the Spring 2001 data on MRI-Plus for the magazines that were covered in both studies reveals that our highest estimates, the ones based on recent reading, generally came much closer to MRI's data in overall magnitude than either recalibrated frequency-of-reading or FRY. For all four versions, our frequency-of-reading-based AIR estimates are all substantially lower than MRI's. For the three versions of the survey with the full complement of 200 magazines, our recent reading percentages for 132 weeklies and monthlies covered by both studies were, on average, within 10 percent of MRI's (based on the average of the ratios of our recent reading figures for individual magazines to MRI's audience figures for those magazines). The category-specific version was the only one whose recent-reading estimates were lower than MRI's, on average. The 80-title version, with the highest six-month reading rates and the highest recent-reading-based read-to-screen ratios, produced much higher audience estimates with recent reading than MRI. This suggests that the smaller number of titles or the prompt to encourage circumspect responses may have unintentionally led to overclaiming of recent reading.

AVERAGE INDEX OF OUR AIR ESTIMATES TO MRI'S	Questionnaire Version			
	Control	Full Deck	Category Filter	80-title[1]
Recalibrated Frequency of Reading	73	71	60	82
Recent Reading[@]	108	107	96	127
First Read Yesterday	63	50	42	59
Number of magazines covered by MRI...				
For Recalibrated Frequency and FRY indices	160	160	160	58
For Recent Reading indices	132	132	132	49

@ Weeklies and monthlies only
Note: *Parade* and *USA Weekend* excluded

[1] See Table A in Appendix 4 for indexes to MRI's estimates for the combination of 80 listed logos and 120 open-end mentions in Version C. See Table B in Appendix 4 for indexes to MRI's estimates, for each of the four versions, based only on the 80 magazines listed explicitly in the 80-title version.

For each one of these four versions, both recent reading and recalibrated frequency-of-reading produce audience estimates that are highly correlated with MRI's AIR estimates. The correlations across magazines for these two methods range from .93 to .98. The shift from the yes/no/not sure form to the check-all-that-apply form for the six-month reading question did not alter the ability of surveys on our platform to capture the underlying variability in reading levels across magazines.

CORRELATIONS WITH MRI AIR ESTIMATES	Questionnaire Version			
	Control	Full-Deck	Category Filter	80-title
Recalibrated Frequency of Reading	0.96	0.96	0.95	0.98
Recent Reading[@]	0.93	0.95	0.95	0.97
First Read Yesterday	0.39	0.71	0.66	0.70

@ Weeklies and monthlies only

IMPLICATIONS

Ironically, we succeeded all too well in one of our original missions – to find a way to add data to our database on six-month reach and frequency-of-reading for 200 magazines that would not lead to inflated AIR figures. Our recalibrated-frequency-of-reading estimates, which were 81% higher than the AIR numbers produced by the industry standard-bearer in our original survey, were 20 to 25% lower in this experiment. It appears that the number of magazines in the first survey and the mode we used to collect the data were not the critical factors in driving the high numbers we were obtaining in that survey. The form of the six-month reading question (i.e., check-all-that-apply with 20 logos per screen vs. yes/no/not sure with six or seven logos per screen), plus alerting the respondent about the number of magazines to be asked about, proved more important.

Our experiment suggests that conducting magazine audience research on the Internet with a random sample can produce credible estimates of average issue reading – estimates that are comparable in magnitude and in variability across magazines to those produced by currently-accepted methodologies. The most viable estimates with our platform, overall, were obtained through recent reading with a pre-coded, non-category-filtered roster of 100-200 magazine logos. Fewer than 100 logos, our test indicates, could lead to inflated AIR estimates, through both higher screen-in rates and higher recent reading rates.

The use of this new medium to collect these data does not imply that they can be produced with radically different approaches. The first-read-yesterday technique, as we implemented it, proved to be problematic, as did our attempts to tap into respondents' latent affinities for magazines by prompting them to think about all of the magazines they read and then list those they had read in the prior six months. Using category filters and grouping titles that fall within the same category, we found, produced lower readership estimates than the other treatments, though this strategy did not alter the closeness of the relationship between those estimates, across magazines, and the magazine audience figures accepted by the industry. For certain categories, particularly in-flight magazines and newspaper supplements, the category-filter/category-grouping option yielded numbers that were closer to MRI's than the ones from the other three versions. Nonetheless, we saw evidence that the categories have to be carefully constructed so that they are specific and recognizable. And we saw evidence that further testing needs to be done to ensure that the researchers' categorization of some magazines matches their readers' perceptions. If these conditions are met, this methodology holds promise for future development as specialized titles proliferate.

Our experiment also succeeded in another of its original purposes -- to craft a single survey over the Web that obtains more information about our panelists' magazine-reading behavior than just their six-month readership and their frequency of reading. We can either solicit this information about double the number of magazines in the original survey (i.e., 200 vs. 100) or solicit this information about the same number of magazines as in the earlier survey (i.e., 100), but also obtain data on newsstand buying, subscriptions, and amount of time spent with a recently-read issue – while producing estimates of magazine reading that do not depart too far from industry estimates. The median length of the "control" interview, which contained all the latter information, was just 12½ minutes.

References

Adelman, David. "Opening Remarks." Transcript Proceedings: Advertising Research Foundatin Magazine Audience Measurement Research Workshop, January 22, 1997.

Baim, J., Arpin, D., and Appel, V. "Using CASI-Audio (Computer-Assisted Self-Interview with Audio) in Readership Measurements." Worldwide Readership Research Symposium, 1995.

Baim, J., Frankel, M., and Arpin, D. "Audio-Visual Computer-Assisted Self-Interviewing (AV-CASI): A Progress Report." Worldwide Readership Research Symposium, 1997.

Beeson, Roger. "Changes in the Great Britain NRS." Worldwide Readership Research Symposium, 1993.

Joyce, Timothy and Coffey, Steve. "Sampling Considerations with Single Source Data." Worldwide Readership Research Symposium, 1993.

Lindberg, Ingemar, "Self-Administered Questionnaires and Frequency of Reading: An Optimal Combination to Measure Readership." Worldwide Readership Research Symposium, 1995.

Meier, Erhard and Finch, Steven. "CAPI and the NRS." Worldwide Readership Research Symposium, 1993.

Napior, David and Brink, Cheryl. "Calibration of Frequency-of-Reading Claims." Worldwide Readership Research Symposium, 1997.

Pinnell, Jon and Appel, Valentine. "How Computerized Interviewing Eliminates the Screen-in Bias of Follow-up Questions." Worldwide Readership Research Symposium, 1995.

Singer, Barbara. "Magazine Media Research 2001." <u>Transcript Proceedings: Advertising Research Foundatin Magazine Audience Measurement Research Workshop</u>, January 22, 1997.

Tchaoussoglou, Costa. "Future Perfect: Uniform Fixed Probabilities?" Worldwide Readership Research Symposium, 1993.

Tessier, Alain J. "The Bigger Picture: Screens, Frequency of Reading, and Readership Claims." Worldwide Research Symposium, 1993.

APPENDIX 1:
EXAMPLE OF SIX-MONTH READING SCREEN IN INITIAL MAGAZINE READING QUESTIONNAIRE

APPENDIX 2:
EXAMPLE OF SIX-MONTH READING SCREEN IN "CONTROL" VERSION OF THE EXPERIMENT

APPENDIX 3:
CLASSIFICATION OF MAGAZINES IN THE CATEGORY-FILTERED VERSION

CATEGORIES IN QUESTIONNAIRE	MAGAZINES CLASSIFIED AS BELONGING TO THOSE CATEGORIES
Automobiles	4 Wheel & Off-Road, Automobile Magazine, Car and Driver, Car Craft, Hot Rod Magazine, Motor Trend, Road and Track
Business or personal finance	Barron's, Black Enterprise, Business Week, E-Company Now, Entrepreneur, Fast Company, Forbes, Fortune, Inc., Individual Investor, Kiplinger's, Money, Mutual Funds, Red Herring, Smart Business for the New Economy, Smart Money, Worth
Computers or technology	Family PC, PC Magazine, PC World, Wired, Yahoo-Internet Life
Women's fashion or beauty magazines	Allure, Cosmopolitan, Elle, Glamour, Harper's Bazaar, Lucky, Mademoiselle, Mode, Victoria, Vogue, W
Food or cooking	Bon Appetit, Cooking Light, Food and Wine, Gourmet, Saveur
Health, fitness, or outdoor activities	Bicycling, Cycle World, Fitness, Health, Muscle & Fitness, Men's Fitness, Men's Health, Outside, Prevention, Runner's World, Self, Shape, Ski, Skiing, Tennis Magazine, Walking, Weight Watchers
Homes, home design, and gardening	Architectural Digest, Better Homes and Gardens, Coastal Living, Country Home, Country Living, Country Living Gardener, Elle Décor, Garden Design, Home, House & Garden, House Beautiful, Martha Stewart Living, Mary Engelbreit's Home Companion, Metropolitan Home, Organic Gardener, Southern Living, Traditional Home
Home remodeling	Family Handyman, This Old House, Today's Homeowner
Hunting and fishing	American Hunter, Bassmaster, Ducks Unlimited, Field & Stream, Guns & Ammo, Hunting, North American Fisherman, North American Hunter, Outdoor Life
In-flight magazines published by airlines	American Way, Hemispheres, Sky Magazine, Southwest Airlines Spirit
Magazines geared to men's interests	Details, Esquire, FHM, Gear, GQ, Maxim, Men's Journal, Penthouse, Playboy, Stuff
Newsmagazines	The Economist, Newsweek, Time, US News & World Report
Magazines that are part of weekend newspapers	Los Angeles Times Magazine, New York Times Magazine, Parade, USA Weekend
Parenting and child care	American Baby, Baby Talk, Child, Family Fun, Family Life, Parenting, Parents, Scholastic Parent and Child, Working Mother
People and personalities	Biography, InStyle, People, Talk, Teen People, Us Weekly, Vanity Fair
Science or nature	Audubon, Discover, National Geographic, Popular Science, Scientific American
Sports	ESPN Magazine, Golf Digest, Golf for Women, Golf Magazine, Golf World, Sport, Sporting News, Sports Afield, Sports Illustrated, USA Today Baseball Weekly
Travel	Conde Nast Traveler, National Geographic Traveler, Travel & Leisure
TV, music, or entertainment	The Cable Guide, Entertainment Weekly, Jet, Movieline, Premiere, Rolling Stone, Soap Opera Digest, Soap Opera Weekly, The Source, Spin, Stereo Review, TV Guide, Vibe
Other magazines geared to women's interests (This category always followed "Women's Fashion or Beauty" in the questionnaire)	Bridal Guide Magazine, Bride's, Cosmo Girl, Essence, Family Circle, First for Women, Good Housekeeping, Jane, Ladies' Home Journal, Marie Claire, Modern Bride, More, O, Redbook, Rosie, Seventeen, Teen, True Story, Woman's Day, Woman's World, Working Woman, YM
Uncategorized magazines: Screen #1	American Photo, Mature Outlook, Midwest Living, Modern Maturity, New Choices, New York Magazine, Popular Mechanics, Popular Photography, Real Simple, Sunset, Texas Monthly
Uncategorized magazines: Screen #2	Atlantic Monthly, Ebony, National Enquirer, New Yorker, Reader's Digest, Smithsonian, The Star, Town & Country

APPENDIX 4:
SUMMARY OF COMPARISONS ACROSS VERSIONS

A. COMPARISONS BASED ON ALL 200 MAGAZINES

	Control*	Full-Deck	Category-Filtered	80-Title: Listed Plus Open-End
Average number of magazines read in prior 6 months	12.1	11.5	10.8	7.0
Average percent read each magazine in prior six months	6.0	5.7	5.4	3.5
Average Read-to-Screen Ratio based on recalibrated frequency of reading	.39	.39	.35	.55
Average Read-to-Screen Ratio based on recent reading	.55	.56	.55	.72
Average Read-to-Screen Ratio based on FRY	.32	.35	.23	Not available
Average AIR based on recalibrated Frequency of Reading[1]	2.3	2.2	1.9	1.4
Average AIR based on Recent Reading[1]	3.6	3.5	3.1	2.3
Average AIR based on FRY[1]	1.9	1.9	1.5	Not available
Average Index to MRI: AIR based on recalibrated frequency of reading[1]	73	71	60	39
Average Index to MRI: AIR based on recent reading[1]	108	107	96	59
Average Index to MRI: AIR based on FRY[1]	63	50	42	Not available

B. COMPARISONS BASED ON 80 MAGAZINES LISTED IN 80-TITLE VERSION

	Control*	Full-Deck	Category-Filtered	80-Title: Listed Titles Only
Average percent read each magazine in prior six months	7.0	6.5	6.2	7.6
Average Read-to-Screen Ratio based on recalibrated frequency of reading	.38	.40	.36	.39
Average Read-to-Screen Ratio based on recent reading	.58	.56	.55	.59
Average Read-to-Screen Ratio based on FRY	.34	.36	.20	.39
Average AIR based on recalibrated Frequency of Reading[1]	2.6	2.5	2.2	2.9
Average AIR based on Recent Reading[1]	4.3	4.0	3.5	4.9
Average AIR based on FRY[1]	2.3	2.2	1.7	2.0
Average Index to MRI: AIR based on recalibrated frequency of reading[1]	72	72	62	82
Average Index to MRI: AIR based on recent reading[1]	114	99	109	127
Average Index to MRI: AIR based on FRY[1]	64	36	41	59

* Combination of two versions with 100 magazines each   [1] Parade and USA Weekend included

Recent reading comparisons based on weeklies and monthlies only

APPENDIX 5:
SIX-MONTH READING PERCENTAGES AND ESTIMATED AIRs, BY CATEGORY

Percent read in past 6 months -- Average for each category

CATEGORY	Control Version	Full-Deck Version	Average of these two versions	Category-Filtered Version	Percent Difference between Category-Filtered Version and Average of other two Versions
Automotive	5.3	6.3	5.8	5.3	-10.2%
Business and personal finance	2.8	2.8	2.8	2.5	-11.9%
Computers and technology	3.6	3.5	3.5	3.1	-14.0%
Fashion and beauty	5.4	4.9	5.1	5.5	7.1%
FOOD AND COOKING	4.4	4.7	4.6	5.5	**16.8%**
Health, fitness, outdoor activities	3.5	3.7	3.6	3.5	-1.9%
Home design and gardening	8.0	6.8	7.4	6.8	-9.7%
Home remodeling	3.0	3.9	3.4	3.9	11.1%
Hunting and fishing	4.6	5.3	5.0	3.3	-51.1%
IN-FLIGHT	1.0	0.4	0.7	1.8	**61.8%**
Men's interests	4.4	4.6	4.5	3.8	-20.4%
Newsmagazines	19.6	18.1	18.9	18.0	-4.9%
NEWSPAPER SUPPLEMENT	4.2	5.1	4.6	10.5	**56.1%**
Parenting and child care	5.4	4.0	4.7	4.4	-6.3%
People and personalities	10.9	10.5	10.7	8.8	-21.3%
Science and nature	8.6	7.6	8.1	6.9	-17.7%
Sports	5.6	5.7	5.7	4.8	-19.0%
Travel	4.2	3.5	3.9	4.7	17.6%
TV, music, or entertainment	7.8	7.3	7.5	6.3	-20.0%
Women's interests	7.6	6.6	7.1	5.4	-32.7%

APPENDIX 5 (Continued):
SIX-MONTH READING PERCENTAGES AND ESTIMATED AIRs, BY CATEGORY

CATEGORY	Percent Reading an Average Issue, as Measured by Recalibrated Frequency of Reading – Avg for each category					Percent reading an average issue, as Measured by Recent Reading -- Average within category (weeklies and monthlies only)				
	Control Version	Full-Deck Version	Average of these two versions	Category-Filtered Version	Percent Difference between Category-Filtered Version and Average of other two Versions	Control Version	Full-Deck Version	Average of these two versions	Category-Filtered Version	Percent Difference between Category-Filtered Version and Average of other two Versions
Automotive	2.1	2.4	2.2	2.0	-11.9%	3.4	4.1	3.7	3.5	-5.7%
Business and personal finance	0.9	0.9	0.9	0.7	-20.1%	1.3	1.2	1.3	1.2	-6.0%
Computers and technology	1.4	1.5	1.5	1.1	-31.4%	1.9	2.0	1.9	1.5	-30.3%
Fashion and beauty	2.0	1.7	1.9	1.8	-3.3%	3.0	2.6	2.8	3.2	11.6%
FOOD AND COOKING	1.7	1.7	1.7	2.0	**14.8%**	3.2	2.9	3.1	3.6	**14.3%**
Health, fitness, outdoor activities	1.4	1.6	1.5	1.4	-8.0%	2.3	2.6	2.5	2.3	-7.9%
Home design and gardening	3.1	2.7	2.9	2.4	-20.7%	7.9	6.4	7.2	6.2	-14.7%
Home remodeling	1.0	1.6	1.3	1.5	13.9%	1.7	2.3	2.0	2.1	2.5%
Hunting and fishing	1.9	2.2	2.1	1.2	-64.6%	3.4	4.3	3.8	2.5	-51.6%
IN-FLIGHT	0.4	0.0	0.2	0.4	**46.9%**	0.6	0.2	0.4	0.7	**46.2%**
Men's interests	1.8	1.9	1.9	1.4	-31.1%	3.0	3.2	3.1	2.4	-28.5%
Newsmagazines	5.0	5.4	5.2	4.8	-8.3%	5.7	6.1	5.9	5.8	-1.7%
NEWSPAPER SUPPLEMENT	1.4	2.0	1.7	4.8	**64.4%**	1.8	2.5	2.2	6.1	**64.8%**
Parenting and child care	2.2	1.6	1.9	1.5	-25.4%	3.2	2.5	2.8	2.8	-0.5%
People and personalities	3.4	3.4	3.4	2.5	-35.4%	5.1	5.3	5.2	4.1	-25.5%
Science and nature	3.6	3.3	3.4	2.7	-24.9%	5.1	5.0	5.1	4.3	-17.8%
Sports	1.8	1.8	1.8	1.5	-23.6%	2.4	2.7	2.5	2.1	-18.7%
Travel	1.5	1.4	1.4	1.5	4.0%	2.5	2.4	2.4	2.7	10.7%
TV, music, or entertainment	3.1	3.0	3.1	2.2	-38.4%	4.1	3.9	4.0	3.2	-26.8%
Women's interests	3.1	2.8	2.9	2.1	-39.3%	4.4	3.8	4.1	3.0	-35.4%