# CALIBRATION OF FREQUENCY-OF-READING SCALES

**Cheryl Brink, PC World, USA**
**David Napior, Roper Starch Worldwide, USA**

## Abstract

An important source of variation in average-issue audience and turn-over rate estimates produced from frequency-of-reading responses is the calibration scheme applied to those responses. AIA estimates derived from stated probabilities are compared to their empirical probability counterparts. Empirical probabilities are shown to be systematically related to publishing frequency and readership loyalty. Calibration implications are drawn.

## Introduction

Many print audience measurement services rely on self-reports of frequency-of-reading of specific titles as the basis for measurement of average issue audience. Virtually all of the other services use frequency-of-reading data for calculation of turnover rates (Douglas and Jones, 1995). This universal dependence on frequency data makes it an important area of study. However, there has been relatively limited discussion of it in the media research literature from either a cognitive psychology or psychometric point of view. Greater understanding is needed with regard to respondent strategies for answering frequency questions, and the implications of those strategies for applying calibration schemes to compute average issue audience estimates from frequency responses.

When forming judgments of frequency-of-reading, respondents have, by and large, two contrasting cognitive strategies available to them:

- episodic-counting (recall of specific, discrete events or behaviors)
- rate-of-occurrence estimation (reporting of a perceived pattern of behavior)

(Blair and Burton, 1987; Burton and Blair, 1991).

An x-of-*last/past*-four line of questioning without a six-month screen (e.g., Monroe-Mendelsohn) evokes an episodic-counting strategy from respondents. An x-out-of-*every/average/usual*-4 line (e.g., IntelliQuest CIMS, Simmons CompPro) evokes a rate-of-occurrence strategy. Those services using an x-of-past-4 question but also employing a six-month screen (e.g., JD Power, Erdos & Morgan Purchase Influence in American Business),) have a somewhat ambiguous assignment to the episodic counting/rate-of-occurrence paradigm, although they seem, on balance, to be more likely to evoke an episodic counting strategy from a respondent then a rate-of-occurrence strategy.

Production of average-issue audience estimates and turn-over rates from frequency-of-reading data requires calibration of the reading probabilities assigned to the frequency-of-reading response categories. Episodic-counting and rate-of-occurrence estimation strategies indicate contrasting approaches to this calibration.

Episodic-counting is best suited to *stated probability* calibration. In stated probability calibration, the frequency response categories are assigned their literal or face-value probability values (i.e., 4-of-4=1.0; 3-of-4=.75; 2-of-4=0.5; 1-of-4=.25, and 0-of-4=0.0). These probabilities have clear merit when applied to a discrete set of potential episodes.

On the other hand, stated probability calibration is not well-suited to rate-of-occurrence responses. Some insight into why this is so can be gained from latent structure measurement theory (Lazarsfeld and Henry, 1980). From the latent structure perspective, the responses to the frequency question would be viewed as observed indicators of latent (true) probabilities of reading an average issue over a period of time much longer than that literally specified in the question, say at least as long as the screen interval (six months in most consumer magazine studies) or perhaps even as long as the inter-study interval, usually one year.

The implications of the latent structure perspective are most apparent at the 0-of-4 and 4-of-4 tails of the frequency response spectrum. For example, let us assume the latent structure of concern to be the probability of reading an average issue of a weekly publication over a six-month, 26-issue time span and that our observed data is based on an x-of-average-four-issue frequency-of-reading question coupled with a six-month screen. Then, in a world of perfectly logical respondents, the 4-of-4 response category would be appropriate for readers of 23, 24, 25, or 26-of-26 issues. The latent (true) probability of reading associated with the 4-of-4 response would be somewhere between .88 and 1.0, depending on the relative frequency distribution of the 23, 24, 25 and 26-of-26 types of readers in the target population. The 1.0 value assigned by the stated probability model to the 4-of-4 response is almost certainly an overestimate of the true probability from the latent structure perspective.

The direction of the skew of stated *vs* latent probabilities is just the opposite at the 0-of-4 end of the frequency response spectrum. Responses to the six-month screen question provide an additional data element relevant to the calibration of the 0-of-4 response. A 0-of-4 respondent who fails the screen will be assigned a latent reading probability of zero, as in the stated probability model; however, a 0-of-4 respondent passing the screen will have a latent (true) probability somewhere between zero and .125; hence, the literal model is almost certain to skew downward relative to the latent structure calibration of the 0-of-4 response category.

The net impact of these countervailing forces at the two ends of the frequency response spectrum will depend on the relative frequency of 0-of-4 and 4-of-4 respondents.

The latent structure theoretic line of reasoning is not, in itself, sufficient for provision of quantitative estimates of reading probabilities. An empirical methodology for estimation of reading probabilities is needed to determine where the latent (true) probabilities lie. One source for empirical data of this type is MRI. In MRI, respondents are asked to provide both rate-of-occurrence estimates (frequency) and episodic recall (recent reading) for a large number of individual titles. Cross tabulations of these two data sets yields conversion rates-- the proportion of screened-in respondents in a particular x-of-4 frequency-of-reading response category who declare recent readership. These conversion rates are, in effect, probabilities of readership. .

As will be shown in this paper, conversion rates correspond closely with what we might expect the reading probabilities to be, based on latent structure theory. At the same time, these empirical data provide important insights with regard to systematic sources of variation in reading probabilities across titles which can be usefully taken into account in computing average issue audience estimates from frequency data.

## Calibration Analyses

In this section, we provide a heuristic examination of conversion rates in the Spring 1997 MRI Doublebase. Various parameterized and title-specific calibration models are evaluated in terms of how well the AIA estimates derived from the calibrated frequency-of-reading responses correspond to the recent reading AIA estimates for the same set of magazines.

The analysis was restricted, on the basis of adequacy of unweighted sample bases, to:

- 23 weeklies
- 6 bi-weeklies
- 2 tri-weeklies
- 109 monthlies
- 15 bi-monthlies.

We begin the analysis by comparing stated probability AIAs with recent reading estimates of AIA. For the stated probability AIA computations, we apply a reading probability of 1.0 for 4-of-4's, 0.75 for 3-0f-4's, 0.5 for 2-of-4's, 0.25 for 1-0f-4's and 0.0 for 0-of-4's.

**Average Percent Difference Between Stated Probability AIA and Recent Reading AIA Overall and by Publishing Frequency**

| | Titles | Average Difference | Standard Deviation |
|---|---|---|---|
| **Total** | 155 | -2.2% | 10.4% |
| **Weeklies** | 23 | 9.1% | 8.8% |
| **Bi-weeklies** | 6 | -4.1% | 7.3% |
| **Tri-Weeklies** | 2 | -7.8% | 2.8% |
| **Monthlies** | 109 | -5.6% | 5.9% |
| **Bi-monthlies** | 15 | -18.6% | 11.1% |

***Source: MRI Spring 1997 Doublebase***

As can be seen, for the MRI data at hand the stated probability AIA estimates are, on average about two percent lower than the recent reading AIA estimates. This relatively small average percent difference rests upon a volatile series of overshots and underestimates, the standard deviation of the title-by-title percent differences amounting to more than ten percentage points.

Furthermore, breakout by publishing frequency of the average percent difference between stated probability AIA relative and to recency AIA yields some interesting regularities. Stated probability AIAs for weeklies are, on average, nine percent higher than the recency AIAs. On the other hand the stated probability AIAs for less frequently published titles tend to be consistently lower than the recency AIAs. This finding from MRI data has been duplicated in other studies in which stated probability AIAs and recency AIAs have been compared. In one such study, utilizing a different methodology (mail) and frequency question than MRI,, the recency-based AIAs for less frequently published titles were almost one-third larger than their frequency-based counterparts (Valentine Appel for IntelliQuest, 1996).

We now will examine the properties of several empirical calibration schemes, using MRI data. Our approach is to compute, on a title-by-title basis, the conversion rates of each of the frequency-of-reading categories to average issue (recent) readers.

These title-specific conversions rates, when used as reading probabilities, exactly reproduce the recent-reading AIAs for those specific titles. Title-specific rates, derived from cover recognition measures, were ultimately what Roper Starch used in their FARMS '96 study as they transited from an episodic, issue-specific measure of AIA to a rate-of-occurrence, frequency-of-reading based AIA.

There remains, however, a need for a parameterized version of the conversion rate data, particularly for new titles for which conversion rate data is not available or perhaps even for building a general purpose calibration model which might be used by all frequency-of-reading based studies as an alternative to the stated probability approach which is almost universally used.

We began the analysis by looking first at overall average conversion rates in the MRI 1997 Doublebase for the set of 155 magazines cited above. These overall conversion rates were what Roper Starch had initially proposed for the FARMS '96 transition study (Joyce and Napior, 1995).

**Overall Conversion Rates**

| 4-of-4 % | 3-of-4 % | 2-of-4 % | 1-of-4 % | 0-of-4 % |
|---|---|---|---|---|
| 91 | 72 | 52 | 29 | 14 |

When these rates were applied uniformly to the frequency-of-reading responses for all 155 titles, the resulting AIA estimates were, on average, 1.6% higher than the recency AIAs and the standard deviation of the percent differences was 12.3%. This performance in terms of reproducibility of recency AIA was roughly comparable to the stated probability AIA performance described above.

In our effort to improve the calibration scheme we introduced two additional pieces of information about the titles into our conversion rate analysis:

- publishing frequency
- reader loyalty .

Previous analyses have shown a systemic relationship between conversion rates and title publishing frequency (McGlathery & Eadie, Berlin Symposium, 1995). We confirm their 1995 findings for the more recent 1997 MRI data. They are presented below.

**Conversion Rates by Publishing Frequency**

| | 4-of-4 % | 3-of-4 % | 2-of-4 % | 1-of-4 % | 0-of-4 % |
|---|---|---|---|---|---|
| **Weeklies** | 91 | 65 | 44 | 25 | 8 |
| **Bi-weeklies** | 91 | 78 | 54 | 28 | 14 |
| **Tri-Weeklies** | 93 | 75 | 55 | 32 | 15 |
| **Monthlies** | 90 | 74 | 54 | 30 | 16 |
| **Bi-monthlies** | 91 | 79 | 63 | 36 | 19 |
| **Eta** | .09 | .54 | .67 | .48 | .60 |

We can see from the above table that conversion rates tend to be substantially lower for weeklies than for less frequently published titles in all frequency response categories except for 4-of-4. Furthermore, in the 2-of-4 through 4-of-4 categories, the weeklies' conversion rates are all lower than their corresponding stated probabilities.

A possible explanation for this finding follows. Let's assume for the sake of argument that a respondent's ability to remember specific reading occurrences is limited to the most recent four months. (This is a reasonable assumption in that magazine readership is generally not an event of great or lasting importance.) During those four months there are four publishing episodes for a monthly and sixteen or seventeen for a weekly. That means the weekly would have 16 or 17 chances to get into the six month screen versus only 4 chances for the monthly. The odds reverse, however, for conversion from 6-month readers (screens) to average issue readers (reads). The lowest probability that a screen will convert to a read for a monthly is one of out four, but for a weekly it is one out of 16 or 17. As a consequence, no matter what a respondent's frequency of reading, their probability of converting to a recent reader is likely to be higher for a monthly. And it is logical that the place where these differences in probabilities would be most evident is at the lower end of the frequency range (0 out of 4, 1 out of 4, etc.), where respondents are likely to have read very few issues during the time period.

When the publishing frequency average conversion rates are appropriately assigned as empirical reading probabilities the average percent difference between the empirical probability AIA and the recency AIA, across all titles, is 1.5% with a standard deviation of 9.0%.

In pursuit of even greater conformity of the reading probability AIAs to recency AIAs, we next examined the relationship between titles' conversion rates and their reader loyalty. For each title we calculated a reader loyalty index as follows:

reader loyalty = (% 4-of-4) + (% 3-of-4) - (% 1-of-4) - (% 0-of-4).

We then formed five equal-frequency quintiles, labeled very high loyalty to very low loyalty. In order to confirm that this behavioral loyalty index was measuring what we intended, we ran a series of correlational analyses between our index and known indicators of loyalty: mean magazine rating (weighted average ranging from "one of my favorites" to "poor"), mean minutes spent reading, and readers per copy (derived from dividing the projected number of average issue (recent) readers by the US

circulation). As expected, we found moderate to strong positive correlations with the mean magazine rating (r=0.33) and the minutes spent reading (r=0.51), and a significant negative correlation with readers per copy (r=-0.36). The multiple r for all three variables as related to loyalty is 0.65.

The readers per copy relationship is consistent with our in-going hypothesis: Magazines with higher 0 out of 4 and 1 out of 4 responses tend to be those with larger pass-along or public place reading. This type of readership is generally motivated by time specific information or entertainment needs. While this readership can be highly valuable to both readers and advertisers (for example, someone reading a review of PC systems before deciding what to buy), it does not necessarily lead to a longer term commitment to the magazine. As a result, these readers tend to be at the lower bounds of the range for a given frequency of reading response, and should therefore have a lower conversion to average issue readers.

We did not find a statistically significant relationship between loyalty and frequency of publication.

The conversion rates for each of the loyalty quintiles are presented below.

**Conversion Rates by Loyalty Quintile**

| | 4-of-4 % | 3-of-4 % | 2-of-4 % | 1-of-4 % | 0-of-4 % |
|---|---|---|---|---|---|
| **Very High Loyalty** | 95 | 75 | 53 | 34 | 12 |
| **High Loyalty** | 92 | 77 | 56 | 32 | 16 |
| **Moderate Loyalty** | 90 | 73 | 52 | 28 | 14 |
| **Low Loyalty** | 89 | 69 | 49 | 26 | 13 |
| **Very Low Loyalty** | 87 | 68 | 50 | 26 | 13 |
| **Eta** | .76 | .40 | .33 | .51 | .23 |

When the quintile-specific conversion rates are used as reading probabilities, the probability-based AIA estimates are, on average across all 155 magazines, 0.9% higher than the recency AIAs, with a standard deviation of the percent differences of 10.3%. This is not quite as strong a relationship as was established for publishing frequency, but the strength of the loyalty measure in the 4-of-4 response (as measured by the *eta* presented at the bottom of each table) appears to complement the publishing frequency relationship which was relatively weak in the 4-of-4 category.

And importantly, from a theoretical standpoint, the addition of the loyalty parameter to the publishing frequency parameter in our reading probability calibration model makes sense. It accounts for some of the remaining variance in reading probabilities across titles within publishing frequency groups. These differences are the result of variation in reader affinity and circulation dynamics (newsstand versus subscriber circulation, in-home versus out-of-home readership, subscription promotion tactics, pass-along readership, etc.). The frequency-of-reading ratings mask these subtle differences because, as explained earlier in this paper, they incorporate rough rate estimates to summarize discrete, episodic reading events.. This problem can introduce significant error into the average issue audience calculation. Adding a loyalty factor to the calibration scheme tailors the probabilities more to the response curve of the individual publication, thereby improving the accuracy of the average issue audience estimate.

Analysis of conversion rates by publishing frequency and loyalty quintiles combined is presented below.

**Conversion Rates by Publishing Frequency and Reader Loyalty**

| | 4-of-4 % | 3-of-4 % | 2-of-4 % | 1-of-4 % | 0-of-4 % |
|---|---|---|---|---|---|
| **Weeklies** | | | | | |
| **Very High Loyalty** | 96 | 72 | 49 | 32 | 4 |
| **High Loyalty** | 88 | 66 | 43 | 22 | 11 |
| **Moderate Loyalty** | 90 | 63 | 43 | 21 | 10 |
| **Low Loyalty** | 88 | 64 | 43 | 23 | 12 |
| **Very Low Loyalty** | 86 | 54 | 37 | 16 | 7 |
| **Bi-weeklies** | | | | | |
| **High Loyalty** | 92 | 79 | 56 | 31 | 17 |
| **Low Loyalty** | 92 | 75 | 55 | 25 | 11 |
| **Very Low Loyalty** | 86 | 76 | 45 | 20 | 8 |
| **Tri-weeklies** | | | | | |
| **High Loyalty** | 94 | 77 | 56 | 32 | 16 |
| **Very Low Loyalty** | 86 | 66 | 46 | 28 | 11 |

| | 4-of-4 % | 3-of-4 % | 2-of-4 % | 1-of-4 % | 0-of-4 % |
|---|---|---|---|---|---|
| **Monthlies** | | | | | |
| **Very High Loyalty** | 94 | 78 | 57 | 35 | 20 |
| **High Loyalty** | 92 | 76 | 56 | 33 | 16 |
| **Moderate Loyalty** | 90 | 75 | 54 | 30 | 15 |
| **Low Loyalty** | 89 | 72 | 51 | 28 | 15 |
| **Very Low Loyalty** | 87 | 69 | 51 | 26 | 13 |
| **Bi-Monthlies** | | | | | |
| **High Loyalty** | 95 | 84 | 64 | 37 | 14 |
| **Moderate Loyalty** | 91 | 92 | 61 | 35 | 14 |
| **Low Loyalty** | 90 | 76 | 63 | 35 | 17 |
| **Very Low Loyalty** | 90 | 78 | 63 | 37 | 21 |
| **Eta** | .81 | .71 | .77 | .78 | .74 |

The two variables, publishing frequency and reader loyalty, when used jointly explain nearly 60% of the variance in the conversion rates. Furthermore, when the publishing frequency/quintile-specific conversion rates are used as reading probabilities, the probability-based AIA estimates are, on average across all 155 magazines, 0.3% higher than the recency AIAs, with a standard deviation of the percent differences of 6.2%.

## Discussion

We have derived a two-parameter reading probability calibration model that works substantially better than stated probabilities in its ability to reproduce a substantial portion of the systematic variation in recency and issue-specific AIAs. The question remains, however, of what utility is such a model.

As we noted above, it has proven to have some value in the Roper Starch FARMS '96 study, which was in transition from issue-specific to a frequency-of-reading based estimation of AIA. For the most part, the FARMS '96 reading probabilities were calibrated according to title-specific conversion rates drawn from three previous waves of the survey.

We found a model similar to the two-parameter model described above to be useful in assigning reading probabilities to previously unmeasured titles and for titles with insufficient sample base to provide stable conversion rate estimates. We anticipate greater use of the model in FARMS '99 as we move further away from our historical benchmark data, and as individual titles' reader loyalty or publishing frequency profiles change from their benchmark status.

Beyond FARMS, we think the calibration model may be useful in recent reading studies as a more stable way to compute empirical probabilities for the frequency-of-reading data also gathered in these studies. These empirical reading probabilities are used in the estimation of a title's turnover rate, one of the two parameters in the beta-binomial audience accumulation model. Title-specific conversion rates are the calibration method of choice and these are generally well-behaved for titles with large base counts, but often display erratic behavior for titles with small base counts. Presumably, application of a smoothing model of the above kind would improve the turnover rate estimates at the unstable end of the title spectrum. For studies with relatively small in-tab samples, most, if not all of the titles, might benefit from this smoothing calibration approach.

To this point, we have discussed internal calibration applications, in the sense that we have concentrated on applications in which the study is using past or current conversion rate data and models to provide current empirical reading probability estimates. What are the prospects for cross-study applications of conversion rate data and models? Unfortunately, not very good. The basic problem is that the relative frequency distribution of the same-title x-of-4 response categories and conversion rates tend to vary substantially from one measurement context to another. In particular, there is a lot of volatility in the differences between data from in-person interviews and mail surveys. Although we can discern no systematic pattern to these differences, we do feel strongly that the volatility of the differences contraindicates application of a reading probability calibration model based on data collected in-person to data collected by self-administered mail questionnaires. The only real hope for more reasonable calibration of frequency-only studies evoking a rate-of-occurrence cognitive response strategy is to add to those studies, on a continuing or periodic basis, an episodic-counting style of question (i.e., recent reading or issue-specific recognition) so that an internal, conversion rate calibration model can be developed.

## References

Appel, Valentine for IntelliQuest, *Two Ways of Measuring Average Issue Readership: Frequency vs. Recency*, November 1996.

Blair, Edward and Scott Burton (1987), "Cognitive Processes Used by Survey Respondents to Answer Behavioral Frequency Questions," *Journal of Consumer Research*, 14 (September), 280-88.

Burton, Scott and Edward Blair (1991), "Task Conditions, Response Formulation Processes, and Response Accuracy for Behavioral Frequency Questions in Surveys," *Public Opinion Quarterly*, 55 (Spring) , 50-79.

Douglas, Stephen A. and Richard D. Jones, "Frequency of Reading for AIA: Under Researched?" *Readership Research Symposium*, Berlin 1995

Joyce, Timothy and David Napior, "Alternative Uses of Magazine Covers as Prompt Aids," *Readership Research Symposium*, Berlin 1995.

Lazarsfeld, Paul F. and Neil W. Henry, *Latent Structure Analysis*, esp. Chapters 6 and 7, "Latent Structure Models with Continuous Latent Space: I and II".