# The Multiple Dimensions of Reader Involvement

Paul Donato
Simmons Market Research Bureau
January, 1991

# The Multiple Dimensions of Reader Involvement

In recent years, the U.S. media research community has begun to actively use qualitative magazine measurements as a means of discriminating between two vehicles which otherwise would have comparable CPMs and compositions. The use of these measures necessitates an understanding of how the instruments used to collect them affect overall levels of reader quality, and, how independent these measures are.

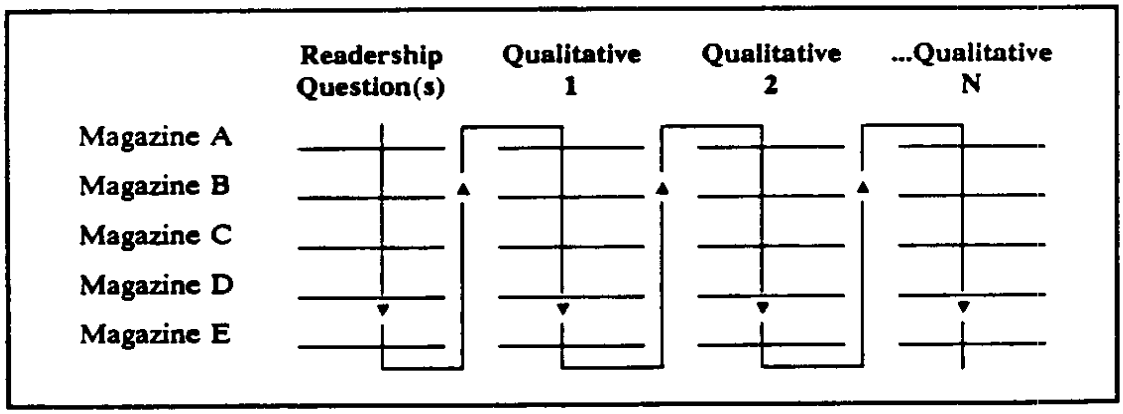
To this end, the Research and Development Group at Simmons Market Research initiated a three stage research program in January, 1989, the objectives of which are:

- to determine the impact of question sequence on reader quality scores developed during personal interviews;
- to evaluate the independence among reader quality measures; and
- to attempt a multivariate reduction of these measures to a more manageable set of summary quality scores.

This paper reports on our findings to date related to the first two stages of this project. The results of the last stage should be available by mid 1991.

## The Impact of Question Sequence

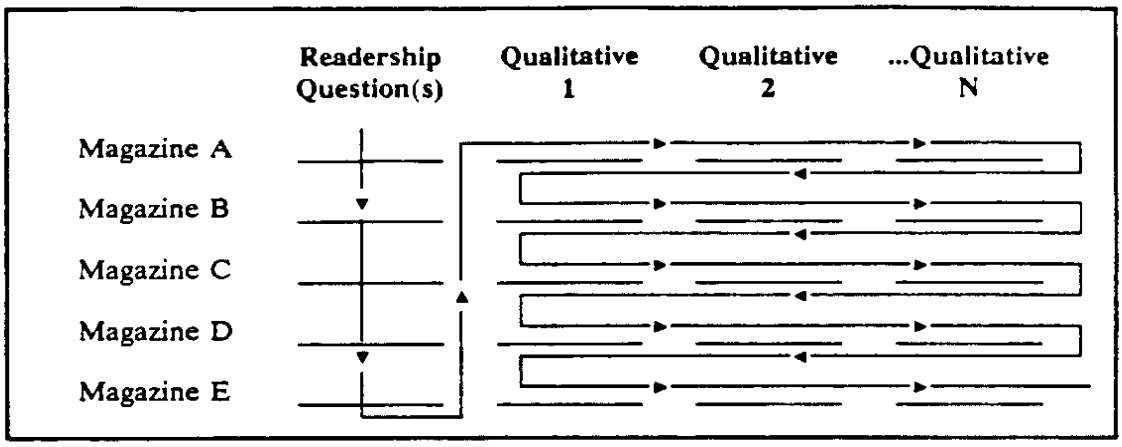
At the outset, it was clear that there are two general sequences with which qualitative questions could be asked. These are illustrated below:




*The Vertical Sequence*

Logic would have that all readership data be recorded prior to asking any qualitative questions so that the burden of answering additional followup questions would not discourage a respondent from acknowledging he/she has read a particular magazine.

However, after readership has been ascertained, the Vertical sequence is characterized by the fact that the first qualitative question is asked of all magazines read (or screened) prior to asking the second qualitative question. This top to bottom sequence continues until all questions have been completed.




*The Horizontal Sequence*

In contrast, using a horizontal sequence, the interviewer first ascertains all readership measures, and then asks the entire battery of qualitative questions for each magazine before going on to the next magazine.

## Test Parameters

The test of sequence impact was conducted between January 22, 1990 and March 16, 1990. A total of 569 interviews were completed. The sample was drawn as an area probability sample of adults 18 years and older residing in the coterminous 48 states. The response rate calculated against the eligible predesignated population was 63%. 301 respondents were questioned using the vertical sequence and 268 respondents were questioned using the horizontal sequence.

| | |
|---|---|
| **Field Dates** | January 1990 - March 1990 |
| **Eligible Predesignated** | 903 |
| **Completed Interviews** | 569 |
| **Vertical Sequence** | 301 |
| **Horizontal Sequence** | 268 |
| **Sample** | Area Probability |
| **Target Population** | Adults 18+ |

*Test Parameters*

The readership measure adopted for the test was recent reading with a six month screen. Recent reading was adopted because the principle concern of the study was qualitative measurement rather than audience measurement and, as a technique, recent reading is far less expensive to conduct.

Nine followup questions were administered after all readership measures had been completed. These were, in order:

- Place of Reading
- # of Reading Days
- Time Spent With Publication
- Actions Taken As A Result of Reading
- How the Last Issue Was Obtained
- % Pages Open
- # of Issues Read Last Day
- Overall Rating
- Interest in Advertising

Of these nine measures, three were excluded from the analysis: *How Last Issue Was Obtained*, *# of Issues Read On Last Reading Day*, and *Actions Taken*. The method of obtaining the *Last Issue* was eliminated from the analysis because there was no implicit way of determining the ordering of responses such as subscriber, newsstand, supermarket. *# of Issues Read On the Last Day* was eliminated because there was minimal variation in respondents' answers. Finally, *Action Taken* was eliminated because there was too little response for all but a few titles.

The analysis was also limited to the 17 magazines which achieved 40 or more respondents.

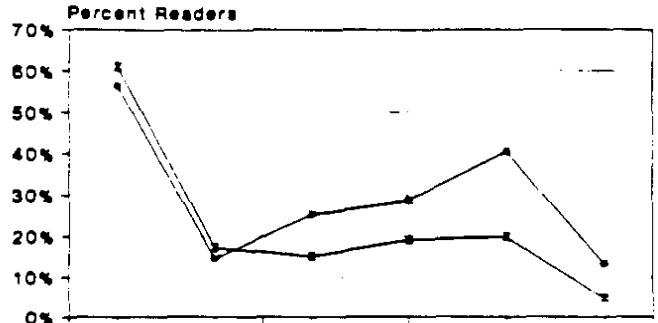
## Findings

In order to compare the results of the two sequences, we computed the percentage of a magazine's readers giving the highest quality answers to each of the qualitative questions. That is, the percent of a magazine's readers:

- reading at home
- reading on 3+ days
- reading one hour or more
- reading 100% of the pages
- citing the magazine as one of their favorites, and,
- expressing considerable interest in advertising.

For example, the following display shows that 61% of Newsweekly (the genre) readers said they read the publication at home when the place of reading question is asked as part of a vertical sequence. 56.3% of Newsweekly readers said they read the publication at home when place of reading was asked as part of a horizonal sequence.

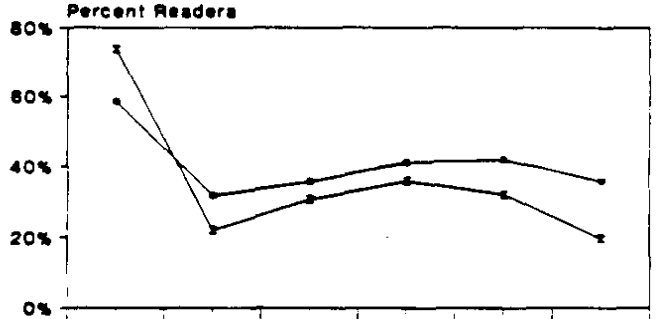
## Newsweeklies




| | Home | 3+ Days | Time | % Pages | Rating | Advint |
|---|---|---|---|---|---|---|
| Vertical | 61% | 17.2% | 15.2% | 19% | 20% | 4.8% |
| Horizontal | 56.3% | 14.6% | 25.2% | 28.6% | 40.5% | 13.4% |

Overall, the display suggests that as one moves further into the question sequence, response quality given under a vertical format declines relative to the horizontal format.

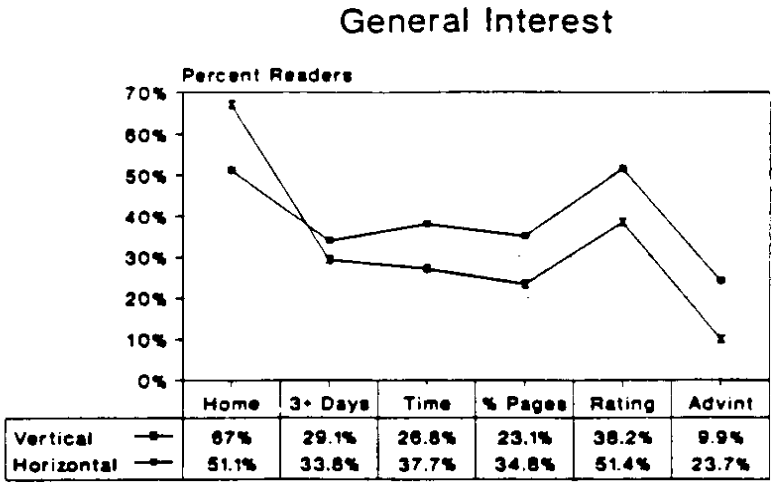
## Women's




| | Home | 3+ Days | Time | % Pages | Rating | Advint |
|---|---|---|---|---|---|---|
| Vertical | 73.4% | 21.9% | 30.6% | 35.6% | 32% | 19.7% |
| Horizontal | 58.4% | 31.6% | 35.7% | 41% | 41.6% | 35.8% |

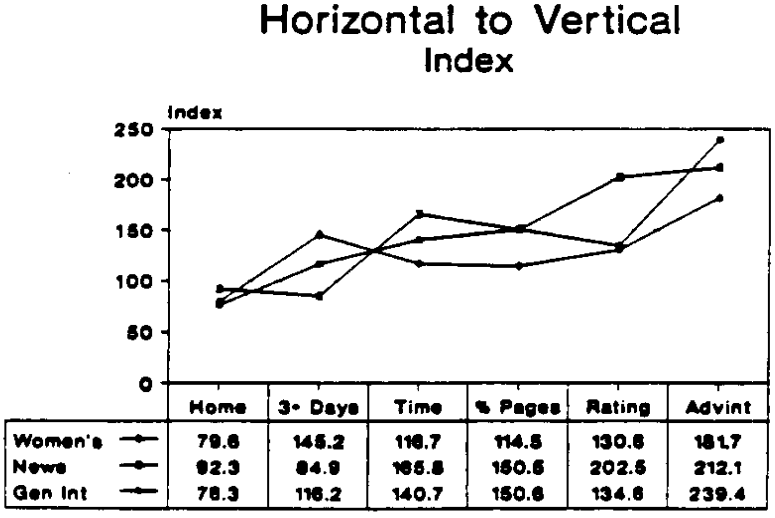
Similar differences in the quality of readership occurred when women's books were analyzed. It is interesting to note that for the first qualitative question, place of reading, respondents exposed to the vertical format were more likely to read at home.

A similar pattern was found for general interest publications, wherein those respondents exposed to the vertical sequence were more likely to read at home than those respondents exposed to the horizontal sequence. However, on the second and all following questions, those respondents exposed to the horizontal sequence yielded higher reader quality scores.




## Interpretation

We believe that these findings provide strong evidence that asking reader quality questions using a vertical sequence yields less positive reader quality scores. Additionally, it appears that the differences between horizontal and vertical scores increase as one moves further into the battery of questions. This is best illustrated by the following display which shows the horizontal score indexed against the vertical score.



For the women's books analyzed, the % at home horizontal/vertical index is 58.4 ÷ 73.4 = 79.6. However, as one moves further into the battery that index increases to 181.7. In fact, for all three genres, there is a clear positive relationship between position in the battery and the horizontal/vertical index.

There is no doubt that the two sequences yield different levels of reader quality. The question is which procedure most properly communicates respondent's attitudes towards print media.

We cannot develop a reasonable hypothesis which would suggest that following a horizontal sequence artificially inflates reader quality responses. However, there is an obvious hypothesis which would explain why questions asked vertically would yield lower reader quality measures: that is the staccato rhythm and lack of focus which results from the vertical format.

This hypothesis has been supported by comments made in a follow up focus group among interviewers who had participated in the test. Interviewers pointed out that by focusing on the magazine (in the horizontal sequence), the battery of questions became a reasonably pleasant opportunity for the respondent to tell the interviewer what they thought of the magazine. "It seemed like they were answering 6 or 7 questions"—one for each magazine they read.

However, focusing on the question (in the vertical sequence), "it seemed like they were answering 50 or 60 questions" (9 questions x 6 magazines read).

The reason for asking qualitative questions vertically is to provide effective comparisons of one magazine to the next before moving on to the next question. Unfortunately, magazines are arranged alphabetically rather than by genre—so qualitative answers are made relative to alphabetical neighbors rather than comparable books within a class. As a consequence, a vertical measurement procedure produces qualitative ratings that may be less meaningful for inter-class comparisons.

## Implications For A Multi-Dimensional Qualitative Program

The implications of question sequence transcend pure methodological issues of instrument design. Rather, the sensitivity to question structure demonstrated in the previous section has much broader implications for our overall attitude towards and use of qualitative questions.

For example, there are, in our opinion as suppliers of these data, two approaches that are frequently expressed or practiced regarding the use of qualitative data:

> the attitude that the instrument affect is less important because it is all relative; and,
>
> the practice of using a single qualitative dimension as a tiebreaker for two books that are otherwise very competitive to the planner.

## "It's All Relative"

The test sample described in the previous section was large enough to realize consistent findings across various types of publications. Unfortunately, the sample was not large enough to compare qualitative rankings of *individual titles* where those ranks would be based on one sequence versus the other. However, there is significant ancillary evidence which suggests that "it's not all relative."

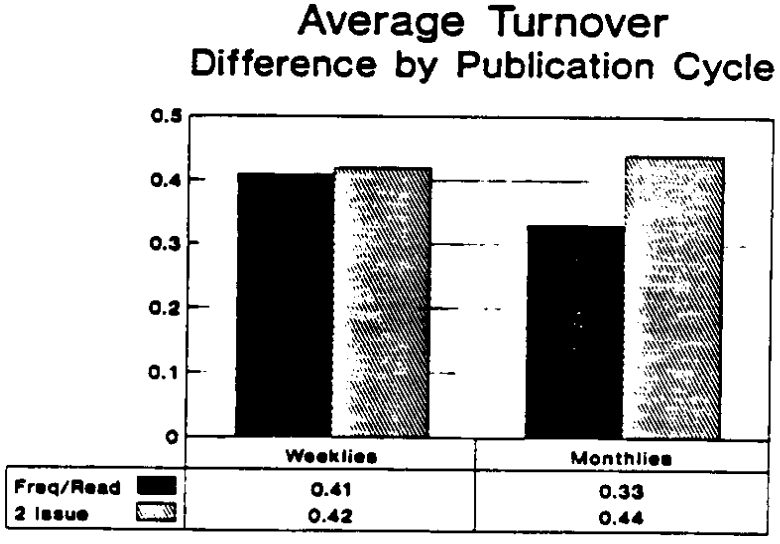
Perhaps the single best example which refutes the hypothesis of "it's all relative" is the case of measuring readership frequency or the loyal reader. There are two common ways to measure frequency (or turnover): empirical—through multiple observations and through a single frequency of reading question (e.g., # of issues read out of 4).

In February, 1989, Simmons, which currently uses two issue-specific observations, conducted a test of 17 publications to determine how turnovers would change if it adopted a frequency of reading question as a means of measuring turnover. The parameters of that test are shown below.

| | |
|---|---|
| Field Dates | Feb. 1989 - March 1989 |
| Eligible Predesignated | 616 |
| Completed Interviews | 425 |
| Response Rate | 69.0% |
| Sample | Area Probability |
| Target Population | Adults 18+ |

*Test of Frequency Measures*

We found that on average, the two issue turnover dropped 17%, from .43 as currently measured with two observations to .36 when measured by a frequency of read (of the last four issues). Most importantly, we found the drop to be less than 3% for weekly publications and 25% for monthlies.



We believe the causal mechanism has been well discussed. If telescoping can affect a difference between recent reading and through the book for a single issue, there is every reason to believe that it may become more problematic as the time frame is extended to four months as in the case of ascertaining readership frequency for a monthly. Respondents are simply more likely to answer three or four out of four issues, relative to the turnover generated by conducting a through-the-book at two points in time.

There is considerable confirming evidence that this difference in the turnovers associated with each instrument exists. For example, in 1989, Simmons conducted a syndicated study of the affluent market in the U.S. using a single interview with a frequency of read question. The following table compares the turnovers (for the same 17 titles shown above) as measured in that study with a frequency question to the same turnovers as measured by two through-the-book interviews in our larger SMM.

| | Two Observations SMM HHI$75K+ | Frequency/Read Affluent Study HHI$75K+ | Δ |
|---|---|---|---|
| Total | .39 | .30 | -23% |
| Weeklies | .38 | .32 | -16% |
| Monthlies | .41 | .28 | -33% |

The point is that monthly publications *appear* to have more loyal readers because of an instrument bias, and therefore, "it's not necessarily all relative." At the same time, this artifical frequency lowers a monthly's capacity to build reach.

## The Use of a Single Qualitative Dimension

In light of this instrument effect, the practice of breaking CPM ties on the basis of a single qualitative dimension becomes considerably more problematic. The previous sections suggest the tie may break differently depending on which instrument was used to develop the adopted qualitative measure.

As important, there is ample evidence to suggest that the tie would break in different directions depending on which qualitative measure was used. In the Fourth International Readership Symposium, Dr. Scott McDonald summarized his own findings and the findings presented at prior conferences:

> Though one would expect a high degree of correlation among all of these measures, in fact, the issue by issue averages are not so perfectly correlated. *(Proceedings, p. 247)*

There is now much evidence that there is only a modest statistical relationship among most qualitative measures.

This, frankly, is to be expected in the social sciences. That is, extreme differences in rates and proportions by groups do not look very significant when expressed as correlations among individual responses.

We illustrate this fact by showing two bivariate qualitative analyses. These analyses are based on a sample of 1,000 early respondents to the 1991 SMM and therefore do not represent the results of a probability sample. The analyses are also limited to four newsweekly publications.

| | | | | |
|---|---|---|---|---|
| | Miss the magazine | Yes | 131 | 56 |
| n = 302 | | No | 31 | 84 |
| $\lambda$ = .50 | | | Yes | No |
| $x^2$ sig. = .0000 | | Reads it as soon as it comes out | | |
| r = .57 | | | | |

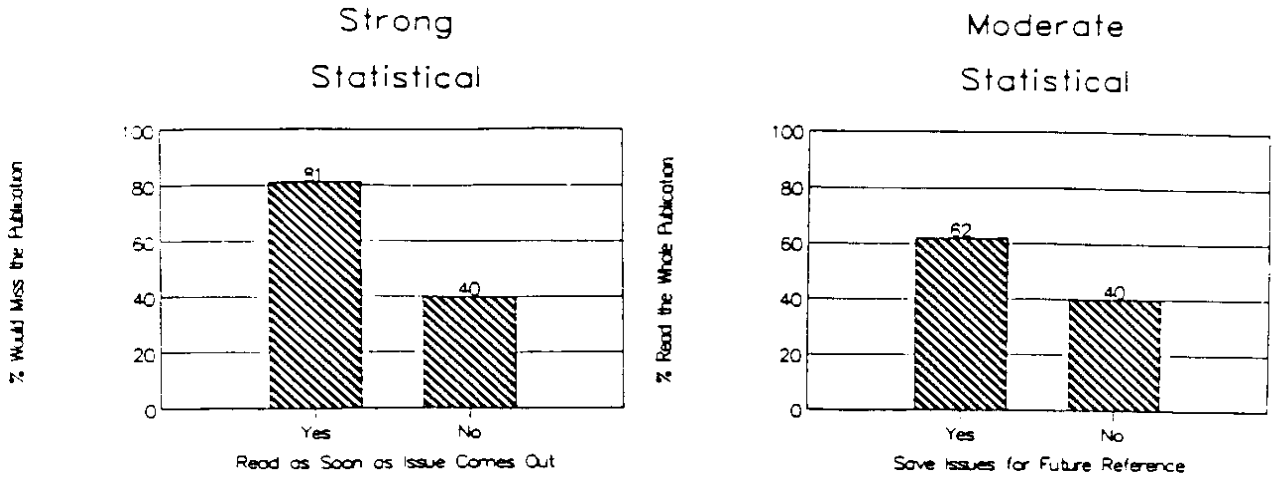
*A Strong Relationship*

81% (131) of the 162 respondents who said they read the publication as soon as it comes out also said they would miss the publication if they could not read it. Only 40% (56) of those 140 not reading it as soon as it comes out, said they would miss the publication. We have reported $\lambda$ as a measure of association because: it makes no assumptions about the nature of the data (nominal, ordinal ...), and, it has conceptual meaning. It is analogous to an $R^2$, and represents the proportional reduction in error (PRE) one makes in predicting one variable once the other variable is known.

| | | | | |
|---|---|---|---|---|
| n = 304 | Generally read the whole thing | yes | 43 | 94 |
| $\lambda$ = .09 | | no | 26 | 141 |
| sig. = .002 | | | yes | no |
| r = .19 | | Save issues for reference | | |

*A Moderate Relationship*

On the other hand, a $\lambda$ of .09 shown above suggests a relatively weak association. However that is because, unless cases land very heavily on the major diagonals, significant prediction error will still occur. Most statistics such as $\lambda$ or r are sensitive to this non-deterministic error. Yet, examine some alternative interpretations of these two tables.




Clearly, people who save issues of newsweeklies are much more likely to read the whole publication and yet an $r^2$ of $.19^2$ or .04 hardly suggests even a moderate association. The point is that looking at $r^2$ or $\lambda$ does not always identify important associations.

These statistical limitations not withstanding, some summary means of evaluating the multiplicity of such relationships must be used. For the reasons already stated, we have adopted $\lambda$ as that mechanism. We extended our analysis of bivariate association to cover nine questions currently measured in the SMM. These are:

- the # of features which appeared interesting as the respondent went through-the-book
- overall opinion
- time spent with publication
- number of readings days
- read as soon as it arrives
- ads are useful source of product information
- save issues for future reference
- I'd miss the magazine if I could not read it
- I generally read the whole issue

$\lambda$'s are shown on both sides of the diagonal to facilitate reading.

## Bivariate Association (λ)
## Among Nine Qualitative Variables

| | interest | opinion | time | days | read/ arrive | ads useful | save | miss | whole |
|---|---|---|---|---|---|---|---|---|---|
| interest | 0 | 0.25 | 0.12 | 0.11 | 0.10 | 0.02 | 0.03 | 0.12 | 0.18 |
| opinion | 0.25 | 0 | 0.17 | 0.07 | 0.16 | 0.0 | 0.01 | 0.02 | 0.04 |
| time | 0.12 | 0.17 | 0 | 0.05 | 0.10 | 0.0 | 0.02 | 0.13 | 0.08 |
| days | 0.11 | 0.07 | 0.05 | 0 | 0.09 | 0.04 | 0.04 | 0.16 | 0.27 |
| read... | 0.10 | 0.16 | 0.1 | 0.09 | 0 | 0.06 | 0.0 | 0.50 | 0.28 |
| ads... | 0.02 | 0.0 | 0.0 | 0.04 | 0.06 | 0 | 0.05 | 0.0 | 0.17 |
| save | 0.03 | 0.01 | 0.02 | 0.04 | 0.0 | 0.05 | 0 | 0.28 | 0.09 |
| miss | 0.12 | 0.02 | 0.13 | 0.16 | 0.5 | 0.0 | 0.28 | 0 | 0.45 |
| whole | 0.18 | 0.04 | 0.08 | 0.27 | 0.28 | 0.17 | 0.09 | 0.45 | 0 |
| avg | 0.12 | 0.09 | 0.08 | 0.10 | 0.16 | 0.04 | 0.07 | 0.21 | 0.20 |

Again, the interpretation is straightforward. A λ of .25 between # of features of interest and overall opinion means that knowing opinion reduces classification error in interest by 25%. An *average* λ of .12 means that knowing the # of features that were of interest reduces classification errors by 12%, on average over the remaining 8 variables.

There are several disparate but relevant conclusions we have reached regarding this table. Some of the more interesting are reported in greater detail.

*Interest in Editorial:* This variable represents the number of articles (of 9) which the respondent identified as of particular interest during the through-the-book interview. Interest in editorial is most closely associated with overall opinion and reading the entire book.

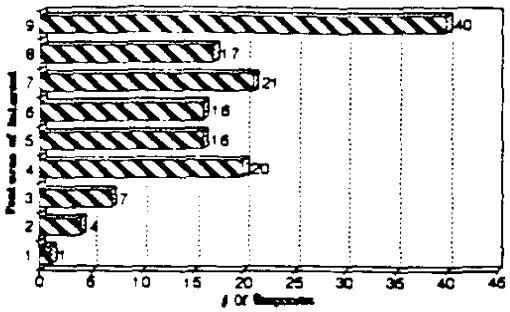
Is a λ of .25 between interest in editorial and overall opinion a strong relationship? Yes, it is as strong as anyone could expect.

| | | Opinion | | | |
|---|---|---|---|---|---|
| | | favorite | good | average or poor | |
| # of Features | 0-4 | 32 (.22) | 57 (.42) | 30 (.75) | $x^2$ sig. = .0000 |
| | 5-6 | 32 (.22) | 53 (.39) | 4 (.10) | λ = .25 |
| | 7-9 | 78 (.55) | 26 (.19) | 5 (.15) | r = -.40 |

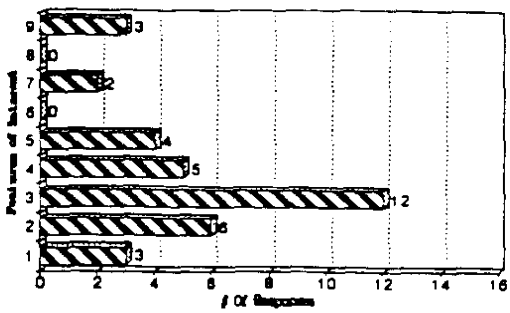
Most (55%) people who consider a magazine one of their favorites, cite 7-9 articles as interesting. 75% of those who thought the newsweekly to be average thought that fewer than 5 articles were of interest. If all cases fell onto the negative diagonal, λ would equal 1.0.

The following displays show that breaking these tables out differently can sometimes show very compelling differences in editorial interest by overall rating of the magazine.

ONE OF MY FAVORITES



AVERAGE



Achieving predictable *differences in rates* by group is all one could expect from such cross-classifications.

*Reading Days and Times Spent:* Most authors have found it curious that time spent and days read are not more closely associated. But whey should they be? The length of the train ride may have more to do with the relationship between these two questions than anything else.

Reading Days

| Time Spent | 1 | 2 | 3 |
|---|---|---|---|
| <30 min. | 48 (.35) | 12 (.14) | 8 (.11) |
| 30 min. - 1 hr. | 50 (.37) | 36 (.42) | 22 (.30) |
| 1 hr.+ | 38 (.28) | 37 (.44) | 43 (.59) |

$x^2$ sig .0000
$\lambda$ = .05
r = .29

However, the data do confirm that most people who pick up a newsweekly on multiple days spend more than 30 minutes with the publication, and that very few people look at a publication over 3 days and spend less than 30 minutes with it. This is logical.

While hardly deterministic, the number of reading days does more accurately predict whether or not someone will read the whole magazine.

| Reading Days | Read Whole Magazine: Yes | Read Whole Magazine: No |
|---|---|---|
| 1 | 42 (.22) | 78 (.51) |
| 2 | 39 (.20) | 49 (.32) |
| 3+ | 112 (.58) | 27 (.18) |

$x^2$ sig = .0000
$\lambda$ = .27
r = -.40

If you pick up a newsweekly on 3 days, you are more likely to read the whole magazine. If you pick it up on fewer than 3 days, you probably don't read the whole magazine. That's a strong relationship even though an $r^2$ suggests only 16% of the variance is explained.

*Advertising As a Useful Source of Product Information:* Perhaps the most important finding in the table is that usefulness of advertising generally has less to do with most parameters of reader quality (avg. $\lambda$ = .04)—at least for newsweeklies. More important is the fact that an analysis of interest in editorial and usefulness of ads proved to be one of the very few analyses which did not achieve statistical significance.

| # of Interesting Features | Ads Useful: Yes | Ads Useful: No |
|---|---|---|
| 1-4 | 39 (.31) | 74 (.42) |
| 5-6 | 41 (.33) | 45 (.25) |
| 7-9 | 46 (.37) | 59 (.33) |

$x^2$ sig. = .146
$\lambda$ = .02
r = .08

This table, then, returns us to the central issue of this last section. What is the consequence of using a single qualitative dimension for the purposes of tiebreaking when CPM rankings are close.

If "usefulness of advertising" is selected and one is evaluating newsweeklies or similar general interest fields, then the irony is that your selection has relatively little to do with magazine quality or reader involvement. In fact, a review of our syndicated database suggests that type of publication is the greatest determinant of advertising interest or utility.

## Conclusions

The paper concludes then, that question sequence affects the overall level of implied reader quality, that there is evidence that instrument biases affect the relative qualitative rankings of magazines, and, that there is sufficient independence among multiple qualitative measures, that selecting one dimension to break a CPM tie will yield results which are a function of the measure selected.

The early affinity models began to address this last conclusion but were criticized for discriminating primarily between types of publications (mass vs. class). These data suggest that simple direct questions may suffer from exactly the same limitations. This points to the need for precise clarity in defining what we want to accomplish with qualitative measures and a multivariate reduction of those measures.

> **If the purpose of qualitative is to measure advertising effectiveness and nothing else, then utility or interest in advertising questions seem to serve the purpose to the extent the respondent can actually provide meaningful data.**
>
> **If, however, we believe the impact of editorial environment is more subliminal—and that people are emotionally drawn to ads placed within interesting copy—then measures of affinity are valuable even though they may not be correlated to advertising interest.**
>
> **If the purpose of qualitative ratings is to measure advertising exposure, then time, days, and proportion of book read or even articles of interest must be measured.**

It is likely that all of these uses of qualitative measures are important and therefore it becomes necessary to reconcile the moderate but logical statistical relationships found among these dimensions of reader quality. These relationships are sometimes modest because some people would really miss a magazine in which they are interested in only half the articles.

Based on these conclusions, this reconciliation of relationships should begin with an objective: defining the factors which are needed in everyday media planning rather than starting with the relationships which exist amqng the variables measured. This suggests that factor analytic techniques used to reduce the data set would be constrained to locate those factors which are required rather than those which occur naturally in the data set.