

MAGAZINE AUDIENCE MEASUREMENT: ITS EVOLUTION AND PITFALLS

THE BEGINNINGS

Although the Audit Bureau of Circulations was created in 1914 to verify circulation claims of subscribing publications, the first ongoing studies of readership didn't begin until the 1930s when Starch began its primary audience studies. As these evolved over the years, a national sample of households was contacted and, with the help of logo recognition aids, adult residents were asked which magazines they or their households subscribed to (or bought at a newsstand) and whether they had read the last issue. Eventually upwards of 60 publications were included, and stability was assured from study to study because Starch "conformed" its findings to known circulation levels for each publication. Hence, if Starch found that 4.5% of the homes in its sample claimed to buy a particular publication, when the "true" figure—based on ABC circulation statements—should have been 4.0%, or 11% lower, all of its primary audience projections were reduced by 11% to bring them in line. Such adjustments rarely exceeded the 8-10% level, either up or down, and more commonly they were in the 2-6% range.

Starch's studies continued into the 1960s, however they presented a static, one dimensional view of magazine audiences, focusing solely on average issue readership, not multiple issue reach and frequency and, of course, on only one segment of the audience—the so-called "primary" reader.

A more sophisticated approach for determining magazine readership paralleled Starch's primary audience studies and eventually caused their demise. In the late-1930s, **Life** began to explore ways to define its total readership, and whether or not such audiences paid for their copies. As a "picture" magazine, **Life** preferred the term "audience" to readers, since this permitted anyone who looked at or scanned its pages to be included, not merely those who "read" the publication. The questioning was tailored accordingly ("Have you read or seen this issue before?").

The early **Life** studies used primitive sampling methods. Initially, interviewers were given daily quotas of adults to contact. Each respondent was shown a full copy of **Life**—page by page—and asked which items, if any, they had seen before. In addition to published issues, issues that had not as yet been circulated were measured as a "confusion control"; invariably, substantial numbers of respondents claimed to have read/seen control issues that they couldn't have been exposed to.

In the early-1940s, **Life**'s research method was modified to speed up the interviewing process and eliminate over-claiming, which was evidently stimulated by the item-by-item questioning.

Continued

Magazine Audience Measurement Continued

As in all future through-the-book studies, respondents were asked to peruse the copy; after this exposure provided the requisite recognition aid, they were asked whether they had read or looked into a copy of that issue any time before the interview. Only those who were certain that this had been the case were considered to be part of the issue's audience.

In 1949 Alfred Politz entered the scene, introducing the probability sampling concept in place of the highly suspect quota samples employed previously. Since this innovation offered the promise of acceptability by advertisers and ad agencies, **Life** initiated a series of custom total audience surveys conducted by the Politz organization. After a benchmark multiple issue study of **Life's** audience in 1950, competitive books such as **The Saturday Evening Post** and **Look** were included in subsequent efforts. Since these visual recognition surveys gave **Life** a substantial reader-per-copy edge over **The Saturday Evening Post**—a “reading book”—and a smaller edge against the bi-weekly **Look**, these publications felt obliged to conduct their own studies also via Politz. Thus, **Look** initiated a nine-magazine study in 1953 that enabled it to compete with both women's service publications and **Life** and **The Saturday Evening Post** for ad dollars (**Look's** constituency was more female-oriented than **Life**). With **The Reader's Digest** accepting advertising in the mid-1950s, a new player entered the game. The **Digest** and the **Post** soon teamed up, sponsoring Politz's “qualitative” reading day and ad page exposure studies, to demonstrate their advantage in this respect over **Life** and **Look**.

Because these were custom studies designed to show their sponsors' strong points as favorably as possible, considerable attention was paid to various forms of validation experiments and, by and large, the findings for the various mass-circulation books measured in one study or another tracked fairly well. The problem from the ad agency/advertiser point of view was that not enough data were provided and far too many publications had no measurements at all.

Responding to the marketing revolution of the late-1950s, ad agencies were increasingly pressed on the accountability issue; data-based media planning was their response. Major agency media departments were, for the first time, asked questions about media mixes, how TV audiences duplicated with print, demographic profiling, budget setting (“how much frequency do we need?”), etc. These, in turn, initiated new inquiries concerning the benefits of various scheduling strategies (“continuity” versus “flighting”) and the comparability of TV, magazine and other forms of media exposures.

In response to these stirrings, Nielsen launched its Nielsen Media Service (NMS) total audience studies in 1960. A dozen mass circulation publications were measured via NMS whose primary selling point was its ability to interface mass magazine audiences—multiple issues were measured allowing 1-4 issue reach and frequency projections—with network TV schedules. (The latter's audiences were derived from diaries kept by the same panel of respondents.)

Continued

Magazine Audience Measurement Continued

Asked by ad agencies to expand its roster of publications, Nielsen proved reluctant, concerned about over-stressing its “field force,” which, in addition to their regular TV meter panel maintenance duties, was also conducting the new magazine studies.

Up to this point, Nielsen’s first study represented the maximum extension of the through-the-book methodology in terms of titles covered (12), and because of this a modification of the basic methodology was necessary. Because of concerns regarding the length of its interviews, the Nielsen Media Service adopted a stratagem not favored by Politz. Although both used a logo screening process to eliminate those who were sure they had not read or seen any issue of each surveyed magazine in the past six months, NMS showed a stripped-down rather than a full copy of the issue to those who screened in as possible or certain readers and therefore qualified for further interviewing. This copy consisted of the cover and table of contents plus 10 or so “representative” editorial items. No ads or recurring columns were included, since these might have been seen in other issues of the same publication or in other magazines. As stated to the respondent, the purpose of this exercise was to determine whether any of the items aroused his or her interest but, in reality, this subterfuge provided the primary visual reminder by exposure to the issue’s key editorial contents. As with the early through-the-book studies, once this perusal was completed, respondents were casually asked whether they had ever read or looked into a copy of that issue before. Those who were certain of prior exposure were counted as part of the issue’s audience; all others were excluded.

While these refinements seemed to work in Nielsen’s mass magazine measurements, which produced findings similar to those of Politz, one assumption that was never adequately validated concerned the age of the issues shown to through-the-book respondents. In the 1950s, Politz had decided that the “optimum” age of copies when shown to a respondent in his recognition studies should be 4-5 weeks for weeklies, 6 weeks for bi-weeklies, and 10-11 weeks for monthlies. In theory, this represented the point beyond which first time readership increments diminished to virtually unmeasurable levels. Also, memory loss and the resulting response bias was felt to be so great as to outweigh the extra effort and cost of extending the age of the copies. Unfortunately, Nielsen and later through-the-book researchers took Politz’s optimum age assumptions as gospel, applying them to numerous small circulation/selective titles never measured by Politz in his 1950s magazine studies.

By the early-1960s, new forces were at play in the advertising and media businesses, and their interactive effects soon produced a revolution in magazine audience research. On the media side, television had in its first decade established itself as the nation’s dominant form of communication, and its ability to deliver mass audiences had made larger and even mid-sized magazines conscious that they lacked the “numbers” to compete effectively. Clearly, total audience—with primary readers swelled by more numerous pass-along audiences—was an

Continued

Magazine Audience Measurement Continued

avenue to competing more effectively with the tube in the numbers game. Moreover, multiple-issue measurements showed a vastly expanded coverage capability for publications who once thought of themselves as reaching only paid circulation constituencies.

Of even greater significance, the deluge of technologically-driven new product innovations following World War II had spawned a new breed of brand managers at the client level who demanded hard evidence to support their decision making process. This led to the rise of copy testing systems and, in the media area, to the end of whimsical TV show sponsorship decisions by the corporate brass. By the early 1960s, the TV networks had largely preempted the advertiser's traditional role as creators/packers of client-sponsored TV shows; the agencies adapted their media activities accordingly. Since they no longer had a partnership-style involvement in TV programs, objectivity now came into vogue, with the agencies delving into demographics, CPMs and other quantitative yardsticks to justify their TV time buys. This quickly spilled over into the new field of media planning and, once they saw a way to position themselves in a leadership role versus rival shops, the large agencies' media departments were encouraged as never before to get "the facts" and use them in all forms of media selection, including print. Quickly BBDO announced a computerized media selection process ("Linear Programming") and just as quickly Young & Rubicam announced its version ("High Assay"); both were followed in short order by other variations, including a multi-agency effort (The "Compass" Model).

Fueling these systems required massive "data banks" of audience and market profile information, so it was not surprising that former Politz sampling expert, Bill Simmons, who had branched out on his own, launched a study of Selective Markets, which won wide acceptance in agency circles. The idea was initiated by the media research director of a large New York ad agency who envisioned a syndicated study eventually involving up to 50-60 magazines, using a large sample base (18,000) and ultimately evolving into a multi-media measurement service including TV, radio and newspapers, plus an expanded marketing/product usage database. Garnering major magazine support, the first Simmons study, conducted in the fall of 1962, provided average issue through-the-book measurements for 36 publications. These included mass books like **Life**, **Look**, **The Saturday Evening Post**, **Better Homes & Gardens**, **Good Housekeeping**, **McCall's** and **The Ladies Home Journal**, all of which had been measured previously by Politz in custom studies. For the first time, Simmons provided total audience data (using stripped-down copies) for large circulation titles like **Time**, **Newsweek**, **U.S. News & World Report**, **Sports Illustrated** and **Woman's Day**; "selective" magazines such as **The New Yorker**, **Saturday Review**, **Esquire**, **Business Week** and **The National Geographic**; and various home service, men's interest (outdoor & science/mechanics), women's fashion and teen titles. Simmons also gathered show-by-show network TV audience data via 24-hour recall questions obtained during the interviewer's visit and, in addition, a limited amount of marketing information from each respondent.

Continued

Magazine Audience Measurement Continued

The whole idea was for an impartial study to cover as many media forms as was reasonably possible, offering information of value to a variety of subscribers, not merely a few sponsoring publications, as in the past. During the mid-1960s, Simmons gradually expanded his magazine roster from 36 to over 50 titles—this being regarded as the limit for through-the-book interviewing, even with the use of the logo card screening process and “skeletonized” issues (10-12 editorial items) to reduce interviewer and respondent fatigue. Meanwhile, a vast expansion in ancillary data gathering occurred. Once respondents had been queried about magazines, selected subsamples were invited to record their TV viewing via personal diaries, and/or to fill out increasingly detailed product/brand usage questionnaires. These were retrieved by Simmons interviewers when they returned for a second round of magazine interviews (whose purpose was to obtain statistical base points for projections of multiple issue reach and frequency patterns). Given a one- and two-issue measurement for each magazine, Simmons could provide extensions for three or more issues, combinations of magazines, TV plus magazine mixes, etc. While the agencies, whose interests went well beyond print audiences, were the primary beneficiaries of the resulting data explosion, magazines quickly learned to position themselves toward agencies and advertisers by demonstrating their reach and frequency capabilities, their ability to target light TV viewers and concentrate exposures on selective product/brand buying segments. In addition, qualitative aspects—for magazines, time spent, reading days and page openings; for TV, viewer attentiveness—were included as basic components of the Simmons studies.

With Simmons gaining unprecedented acceptance, potential rivals recognized that there was money to be made in the syndicated media/marketing measurement field. And so the competitive era began. In 1963, Nielsen belatedly expanded its magazine list to compete more effectively with Simmons, but lost control of its study and in 1964 abandoned its NMS venture (evidence suggested that some members of its overburdened field force began to fill out the magazine questionnaires themselves, erring on the high side in their “readership” claims). The Politz organization, which had turned down the opportunity given to Simmons by the initiating ad agency, entered the field in the mid-1960s. Contending that fuller issues were “better” (producing 8-10% higher readership claims than Simmons’ skeletonized copies), Politz had little new to offer and never mustered much support beyond a few magazine backers. A more ambitious (multi-media) SRDS initiative also was attempted in the mid-1960s, but proved redundant and failed.

The single truly “new” methodology of the 1960s was devised by Norton Garfinkle for his Brand Rating Index (BRI) in 1963. Looking for a cheap method to obtain brand buying/product use information, cross roughed with magazine readership and network TV viewing data, BRI introduced what was essentially an unaided frequency of reading recall study. With the cooperation of respondents, interviewers placed self-administered questionnaires containing numerous product use/brand preference questions for a variety of packaged

Continued

Magazine Audience Measurement Continued

goods and “durables” categories. After many of these questions had been dealt with in the questionnaire, respondents came across a list of magazines that were segregated by issue cycle. In each case, they were asked to indicate whether a) they never read or looked into it, b) they saw it now and then, but had not seen any of the past four issues, or c) they had read either one, two, three or four of the past four issues. The replies to these last four inquiries (read one, two, etc. issues) were weighted to produce an average issue estimate. Thus, someone who had claimed to have seen one issue out of four was given a .25 chance to be an average issue reader, a claimed two-issue reader got a weight of .50, etc. The resulting weighted average produced BRI’s average issue audience estimate.

Many agencies were skeptical about so primitive a methodology, and one media research director who took the trouble to review BRI’s network TV “ratings” obtained in the same manner (“how many of the last four telecasts did you watch?”), found them to be 50-100% higher than corresponding Nielsen meter-/diary-based viewing projections. This was pointed out to BRI and, as if by magic, its TV viewing levels “adjusted themselves” in the next survey to the point where they were more in line with Nielsen’s. This curious development raised questions about the veracity of BRI’s magazine findings. Were they also “adjusted” to conform them to previous Politz, Nielsen and Simmons readership levels?

Eventually, with Simmons still triumphant (having fended off Nielsen, Politz and SRDS), BRI switched to the accepted through-the-book method as a competitive last resort but, like Nielsen, lost control of its 1970 study, aborted the survey and went out of business.

ENTER “RECENT READING”

The BRI experience failed to provide a clearly defined alternative to the through-the-book method as practiced by Politz, Nielsen and most frequently by Simmons. Buried in a self-administered questionnaire, and without the benefit of any visual reminder or interviewer stimulus, BRI’s findings, even if truly reported, were never taken seriously as absolute quantitative measurements. However, BRI’s downfall set the stage for the introduction of yet another competitor to Simmons: Timothy Joyce’s Target Group Index (TGI). Developed in Great Britain, this method had the same advantages as BRI, namely the ability to obtain readership data on a great number of publications at relatively low cost. Launched in 1972, the U.S. version featured an even larger self-administered questionnaire than BRI had used, and once again respondents encountered batteries of media exposure questions deep in the booklet. The core assumption in TGI’s “recent reading” methodology was that claimed reading of any issue (or issues) of a publication, within a time frame proportionate to its issue cycle, produced the equivalent of an average issue audience measurement. For a weekly, persons who claimed any issue exposure in the past week might have seen either the latest and/or one or more older issues. Since issue audiences overlapped as one edition supplanted the other, this form of time

Continued

Magazine Audience Measurement Continued

frame audience determination was believed to cancel out these duplicitous effects, generating a supposedly valid approximation of a single issue's audience as if it had been measured across a number of weeks. Similar assumptions applied to bi-weeklies and monthlies; in the latter case, claimed "any issue" readership within the past month was assumed to render a valid average issue measurement.

The TGI design represented a significant departure from through-the-book, not only in its manner of determining average issue audiences, but also by its inclusion of two forms of audience questions in the same questioning instrument. Respondents were asked how many (if any) of a typical sequence of four issues of the magazine they read—calling for a general "reading pattern" estimate—and, following this, they were questioned about past week or past month exposure to any issue of the publication depending on its issue cycle. By design, this tandem questioning procedure satisfied the demands of ad agencies and publishers for multiple issue reach and frequency projections and, to accommodate them, TGI chose the simplest possible method—asking respondents both forms of inquiries in the same questionnaire, without apparent concern about possible confusion or inflationary biases that might ensue in the readership claims.

Under similar pressure to supply reach and frequency data, Simmons held to the far more expensive option of reinterviewing his respondents about a second issue of each magazine. After the first through-the-book interview, various subsets of the Simmons sample were asked to fill out self-administered marketing and/or TV viewing diaries. These were retrieved by the interviewer in a second visit when, once again, respondents went through the entire through-the-book process for later issues of the publications measured in the first interview. By determining how many out of two issues his respondents had seen for all of the magazines in his study, Simmons could project multiple issue audiences for each title up to full yearly schedules, as well as duplication patterns for combinations of titles. The price he paid was a vastly more expensive study, coupled with the imposition of a heavy additional burden on his interviewers and respondents.

Despite such refinements, the focus of most syndicated audience measurement usage rested on average issue readership projections and their impact on demographic targeting and CPMs. Though its magazine questions were placed well into a bulky marketing and media usage questionnaire, and there was no interviewer to direct the respondent nor visual aids such as logos to aid in recall, TGI produced audience estimates that were 20-25% higher than corresponding Simmons measurements for the same books. Considering that TGI measured more magazines than Simmons, which should have deflated TGI's readership claims, its higher audience levels raised eyebrows. However, now it was Simmons' turn to lose control of one of his studies, inadvertently giving TGI a welcome competitive boost. When Simmons' 1973

Continued

Magazine Audience Measurement Continued

study produced anomalous results, the study was aborted. This vacuum was eagerly filled by TGI, which survived to battle Simmons for the next five years. By measuring so many more magazines and selling its study at cheaper prices, TGI carved out a third of the syndicated magazine audience measurement business, thereby preventing Simmons from making significant profits. By now, Simmons' through-the-book methodology was proving extremely cumbersome as well as costly. Not only did interviewer "kits" require frequent updating with new and properly stripped-down copies of magazines replaced constantly, but the gradual addition of titles (purely for business reasons) and the adoption of year-round interviews produced an ever-mounting strain on the interviewing staff.

TWO METHODS IN CONFLICT

Still, TGI was losing money, and eventually TGI's parent, J. Walter Thompson, threw in the towel; the result was a merger of Simmons and TGI to form SMRB in 1978. This left the new company with a dilemma. How could it retain its traditional through-the-book method—now dangerously over-extended to 70-80 titles—while not losing the business of 40-50 additional books covered solely by TGI? If SMRB didn't do something to woo these customers, it ran the risk of encouraging a new competitive service that would not only fill this vacuum, but also vie with SMRB for the customers it already had.

In response, SMRB came up with a dubious "business" decision. It opted to survey only 40-50 of the larger circulation magazines by through-the-book, while 100 others (including 30 or so once measured via through-the-book) were "demoted" to an "improved" version of TGI's recent reading method. The latter now involved a personal interview with a logo card screening procedure at the outset, followed by questions about the number of issues usually read by the respondent (as noted earlier, such data were needed so the researchers could project multiple-issue reach and frequency patterns). Finally, the recent reading questions (past week or past month readership) for weeklies and monthlies were posed.

Further complicating matters, the competitive vacuum formed by the TGI and Simmons merger was quickly filled by Timothy Joyce's sudden reappearance at the head of Magazine Research, Inc., later changed to Mediamark Research, Inc. (MRI) in 1979. His new—and also improved—recent reading methodology closely resembled SMRB's updated version, using personal interviews, logo card sorts, etc., and once again raised the specter of a competitor able to measure many more magazines than through-the-book at lower prices.

Irrked by SMRB's arbitrary "mixed methodology" proposal, the advertising and publishing industry was galvanized to support what became the Advertising Research Foundation's (ARF) "Comparability Study." Conducted between June and November 1979 (curiously by SMRB), the ARF study involved 4,600 interviews that contrasted the results of three method-

Continued

Magazine Audience Measurement Continued

ologies: a) the now-standard through-the-book recognition technique with a six-month logo screen and skeletonized copies, b) the proposed new recent reading method conducted via personal interviews (rather than in self-administered marketing/media questionnaires), and c) a combination of the two, where respondents were queried on some books via through-the-book while other magazines were measured via the “improved” recent reading methodology.

Unfortunately, the ARF did not explore the crucial issue of validity. Instead, it focused solely on the differences produced by the various methods, and these clearly were substantial. To begin with, the ARF reported that recent reading produced significantly higher (+15%) screen-in rates for monthlies but little difference for weeklies (+2%) relative to through-the-book. Moreover, the recent reading method produced much higher screen-to-read ratios. Whereas the through-the-book interviews found that roughly 40% of their screen-ins subsequently claimed average issue readership for both weeklies *and* monthlies, recent reading noted a 49% screen-to-read rate for weeklies and 63% for monthlies. These critical differences combined to produce 27% and 86% higher audience levels for weeklies and monthlies, respectively, via the recent reading method.

The ARF analysis also noted that the greatest differentials were recent reading’s higher screen-in and reader claims among “opposite sex” audiences (men for women’s books; women for men’s books), infrequent readers and out-of-home readers. These were the primary points of departure; otherwise, the demographic profiles for most publications were, for the most part, similar.

In the face of these findings, SMRB reverted to through-the-book for all publications it would measure and for the next 15 years (1980-1994) this method was pitted against MRI’s recent reading, with the latter company steadily gaining ascendancy. The ad agencies gradually came to the conclusion that despite the differences in reading levels reported by SMRB and MRI, the resulting rank order CPM correlations were about the same for most of the larger- and mid-sized magazines. Thus, with MRI measuring many more books and charging less for its data, its edge over SMRB became insurmountable. Although SMRB increased the number of magazines it measured to well over 100, MRI could always accommodate 60-70 more. To curtail their costs, both services froze their sample sizes at about 20,000 per year even though the addition of numerous small circulation titles pleaded for larger samples. SMRB also cut corners in its stripping-down process and other areas as it tottered under the burden of an increasingly oppressive through-the-book measurement system.

Operating under such stresses, the quality of SMRB’s studies began to suffer. In the early- and mid-1990s, SMRB’s screen-in and screen-to read rates bounced suspiciously, and the Magazine Publishers Association (MPA) raised serious questions about the validity of its

Continued

Magazine Audience Measurement Continued

studies. In 1995, SMRB finally conceded defeat, switching to a “streamlined” version of recent reading that produced 10-15% higher readership levels than MRI. This decision cost SMRB most of its remaining support, leaving MRI in a dominant share-of-market position, which it maintains today.

Holding the relative strengths and weaknesses of the competing methodologies aside, one of the major contributors to SMRB’s demise and MRI’s ascendancy was the changing research orientation and interests of their customer base. The impact of computers, first evident in the 1970s, permeated the advertising and media business, creating a new, electronically-oriented type of researcher. Increasingly, such people were concerned with obtaining “raw” data that could be readily accessed and plugged into their own analytical systems, then tabulated and sent to “end users” to aid their decision-making process. Manipulating masses of data took precedence over interpreting the findings or challenging their validity, and this dovetailed perfectly with the fast paced, numbers-driven media sales and buying mindsets that replaced the more sedate and judgmental practices of the past. With ad agencies downsizing their media planning and media research staffs relative to increased billings, and vast numbers of new media clamoring for evaluation, rapidly available quantitative data provided the most relevant answers to the majority of their media questions. As a result, qualitative factors fell largely by the wayside and few, if any, of the new breed of media sellers, buyers, planners or researchers had the incentive or the training to delve into such murky issues as ad exposure, ad impact, the benefits of media environments, etc.

The syndicated research services contributed to the creation of a data processing rather than a data evaluation mentality among their customers by allowing virtually unlimited access to “raw” information from their studies. In consequence, an agency media planner whose client’s marketing target was women aged 18-34 in the professional/managerial occupations might specify such a demographic “cell” as the basis for audience and CPM analysis, even if the sample base in the latest MRI or SMRB study was only 1,000, not the whole survey’s 20,000 respondents. The computers merrily churned out the audience and CPM projections for every magazine or TV show being evaluated, drawing no distinctions as to accuracy or reliability between data based on 20, 100 or 200 readers.

The overall sample sizes used by the syndicated services should have been a major concern, but this too was overlooked by the agencies. In the mid-1960s, the dominant magazine audience counter, Simmons, interviewed 18,000 adults to measure 50-60 magazines with average issue circulations of 3-3.5 million copies. By the mid-1990s, MRI, the leading syndicated service, was using 20,000 interviews to measure 175+ titles, with an average circulation of 1.5-2.0 million. Clearly, the refusal of the services to increase their sample sizes was governed by rising interviewing costs and the profit motive; just as clearly, this decision

Continued

Magazine Audience Measurement Continued

impacted negatively on the stability of their data, especially when broken down into selective demographic, psychographic or product use segments.

Recently, the magazine industry has evidenced rising concerns about the validity of the recent reading methodology, chiefly among publishers of difficult-to-measure titles, as well as those with more selective constituencies. A common complaint is survey to survey “instabilities” which produce adult reader-per-copy variations that average 8-10% for all magazines on a year to year basis and considerably higher “bounces” for small circulation books. It goes without saying that even larger variations are evident by demographic cells or product use segments, depending on their size.

OTHER DEVELOPMENTS

Several other developments were noteworthy in the readership arena. Although SMRB’s through-the-book, and then MRI’s recent reading interviews provided most of the publishing industry’s mass market audience data in the 1980s and early-1990s, a number of market segments required larger samples and different approaches to measure difficult to interview respondents. J. D. Power and Associates launched its new car buyer readership studies based on Polk registration lists, Monroe/Mendelsohn (now Mendelsohn Media Research) fielded its study of affluent adults, and IntelliQuest launched studies of the computer market. In all three cases, mail questionnaires were employed utilizing frequency of reading rather than recent reading questions. Attitudinal and marketing inquiries that were of unique interest to these industries were also included, while highly selective magazines that were too narrowly based to be handled by the SMRB or MRI mass market studies were also accommodated.

Also of interest was Nielsen’s Home*Scan experience. In the early-1990s, Nielsen attempted to launch a syndicated magazine audience service interfaced with its household panel, which monitored packaged goods purchases with hand-held electronic scanners and TV set usage via a meter device. Designed by Timothy Joyce, who had left MRI, Home*Scan used a recent reading methodology which supplied instructions and materials (card decks, answer boards, etc.) to the respondent for a self-administered survey. The familiar card sort screening process was followed by frequency of reading and recent reading questions, and all answers were recorded by specially adapted scanners. Approximately 170 publications were measured in Home*Scan’s first effort, which produced readership claims that were 44% higher than MRI’s personal interview-based recent reading findings and 86% higher than SMRB’s through-the-book interviews. Although various rationales were cited to explain these differences (Nielsen used a panel, not a probability sample, so its respondents may have over-cooperated or been atypical, demographically), the most likely culprit was the self-administered aspect and lack of interviewer control. Unable to defend its unacceptably high readership levels, Nielsen dropped Home*Scan after conducting several studies.

Continued

Magazine Audience Measurement Continued

OBSERVATIONS & CONCLUSIONS

Throughout its evolution, the purity of magazine audience research designs and the quality of their execution has been influenced by combinations of “external” and “internal” factors. By “external,” we refer to the competitive posturing and gamesmanship between rival syndicated services, plus the varying attitudes and needs of their customers. By “internal” we mean the professional skill and the ego gratification drives of research designers, their ability to control and execute studies according to plan, and finally, their perfectly understandable preoccupation with the profit motive. Needless to say, these variables are intertwined. When two services are battling it out for supremacy, it is only natural for each to covet the other’s exclusive customer base, particularly in a situation where profits are slim or non-existent. By using the cheaper-to-execute recent reading method, first TGI, and then MRI, were able to compete against a supposedly “better” through-the-book system by the simple expedient of covering many more magazines and charging less for their data. Not surprisingly, this led Simmons and its successor, SMRB, to add more and more titles to their studies, even though this should have raised concerns about overburdening its interviewers and fatiguing its respondents. Similarly, as both SMRB and MRI added smaller circulation magazines, the obvious need for larger samples was subordinated to the quest for profitability.

In the 1960s, such a situation would not have been tolerated by the ad agencies who, at that time, were greatly concerned about the accuracy of the through-the-book findings. However, in the “data for data’s sake” mindset that took root in the mid-1970s and has since entrenched itself, the agencies stopped acting as the guardians of validity. Instead, they demanded readership statistics for as many books as possible, while “simulated” readership estimates were made when surveys for new publications weren’t available. This change in focus is the primary explanation for the lack of agency pressure on SMRB and MRI over methodological and sample size issues since the late-1970s. As noted earlier, unlimited data access has also contributed to the problem. When Bill Simmons allowed magazines and ad agencies to acquire what amounted to raw data tapes in the early-1970s, he saw this as a way to ingratiate himself with his customer base. His competitors, of course, followed suit, and even though the contending parties belatedly recognized that the integrity of their studies was being abused by undisciplined fine cell tabulations, while a host of otherwise salable marketing information was given away free of charge, it was too late to redress the situation. If SMRB abruptly stopped issuing tapes while MRI continued this practice, SMRB might lose customers. As a result, the data was allowed to flow unchecked and uncontrolled. Worse, whenever a customer processed the information without regard for common sense sample size limitations and then found the data to be “unstable” or “erratic,” the blame fell on the research service.

As for the validity of the research designs, it goes without saying that basic business pressures and customer demands have had a paramount and often detrimental effect. Although the

Continued

Magazine Audience Measurement Continued

syndicated services were supposed to be impartial, the fact remained that magazines contributed the lion's share of their revenues. As the agencies' critical scrutiny waned, the services became increasingly sensitive to the promotional needs of the publications. This prodded them to develop methods that produced higher readership claims, hence the long succession of "improved" recent reading designs, each raising audience levels in a quest to avoid "understatement."

Finally, we should note the magazine and advertising industries' collective failure to commission a definitive validation study that would determine which method and execution procedure was the most accurate way to measure readership. For a variety of "political" reasons, the definition of "readership" has never been reviewed seriously, the main cause being publishers' fears that this might lead to a reduction of reported audience levels that, in turn, might position magazines in a negative light vis-a-vis TV. And last but not least, certain types of publications ("fast reads," pictorial formats, etc.) benefit to a far greater extent than others from the maximized total audience (read/looked into) readership definition. Why should these publishers give their opponents (reading-intensive magazines) a competitive sales benefit by exploring this question?

Even if we hold economic pressures aside, the issues a researcher must deal with in designing magazine audience research are daunting. Most important is the questioning method. Should it be a personal interview or can it be done by phone or by mail? In addition to the basic cost factor which impacts heavily on these and attendant sample size decisions, these are some of the factors to be considered:

1. The nature and number of titles to be measured.
2. How much ancillary information (multiple-issue reach & frequency data, marketing information, TV viewing estimates, etc.) is to be collected.
3. Whether the universe to be sampled is the mass population or a difficult-to-interview segment.

Even when these issues are resolved, the research designer must determine the clearest and most objective way to frame the question sequence—heightening respondent cooperation, while minimizing response bias. The latter involves such issues as memory loss, title confusion and prestige claims. Another matter to be considered is whether the respondent should know from the outset that this is an audience study—perhaps causing over-claiming to impress the interviewer—or should the real purpose be disguised, as in through-the-book, in an effort to obtain more factual responses.

Continued

Magazine Audience Measurement Continued

To show how these variables interact, let's consider the way hypothetical through-the-book and recent reading interviews would progress as the reader of this report might be subjected to them. First, let's take a through-the-book interview in which a reasonable number of publications (60) are measured. Put yourself in the typical respondent's place.

An interviewer arrives at your home, secures your cooperation for an opinion study and after a few preliminary questions pulls out a 20-page looseleaf booklet. Each page contains three four-color magazine logos. As these are displayed, you are asked whether you have read or looked into any issue of each publication in the past six months. In every instance, the interviewer asks whether you are sure about your answer and notes your replies. At this point you aren't really aware that you are participating in an audience study. Also, with the interviewer asking about the certainty of your replies, you rethink your answers in a few cases, moving one magazine from *sure* to *may have read* status and, on reflection, moving another from the *maybe* to *sure not read* category. In both instances, the interviewer accepts and records these changes.

All told, you identify yourself as being a *certain* past six months reader of three magazines and a *possible* reader of three others. This triggers an "editorial interest" interview for each of these six publications. A skeletonized copy of a five-week old issue of a weekly magazine is produced. It contains the cover, a table of contents and ten editorial items unique to that issue. As you scan the issue, the interviewer asks whether each item is interesting and notes your replies. Then, as if this is just a formality, the interviewer asks whether you have ever read or looked into a copy of this particular issue before, including any away-from-home exposures, regardless of whether you bought the copy, subscribed to it or obtained in some other way.

This, of course, is the key total audience question, and here too, the interviewer tries to establish certainty. You don't know it, but only those who are sure they have perused the magazine will be counted as "readers."

Recognizing several of the articles and the cover, you answer that you have seen a copy of this issue before. And yes, you are sure of it.

Before moving on, the interviewer asks a number of other questions about the magazine you just claimed to have read. These include how much time you spent reading the issue, how many days you read it, where this occurred, and whether you took any of a specified list of actions (clipped coupons, tore out articles, etc.).

As the interviewer picks up the next magazine for your perusal, you vaguely remember that only 50 or 60, or perhaps 70 logos were in the booklet, which seems to serve as the starting point for these more detailed questioning sequences. You ask yourself how much more time

Continued

Magazine Audience Measurement Continued

will be spent with the interviewer, who now wants you to examine a copy of a monthly (first published 10 weeks ago). You remember her claim that the process won't take long and continue. After examining this and four more stripped-down monthly magazines, with all the attendant questions described above, the "audience" phase of your interview concludes. The whole process took 35-40 minutes. While you had trouble recognizing several of the magazines, you did the best you could under the circumstances and you feel generally confident about your replies. The interviewer, who seemed fairly competent, agrees and thanks you for cooperating.

In contrast, let's compare the reasonably straightforward 1960s-style through-the-book procedure, described above, with a hypothetical recent reading interview.

Once again, you are asked to sit for what is described as a short interview. The interviewer poses a few questions about other matters, and then pulls out a deck of yellow cards—there appear to be about 50 of them (actually there are 40). As these are shuffled, the interviewer explains that each card depicts the title and logo (in color) of a different magazine. Next, the interviewer opens what appears to be an answer board on which are printed responses to various questions.

Now the interviewer starts handing you the cards one by one. In each case, you are supposed to put the card in an appropriate place on the board, depending on your answer. The question for now is, "Have you read or looked into this magazine in the past six months?" There are three response options, represented as boxes, on the board. One is for a yes with certainty (*sure*) answer, another is for *not sure* and the third box is for a definite *no*. Before you begin, the interviewer reminds you to consider any exposure at or away from home, no matter how you got the copy, or how thoroughly you read/looked into it. Obviously this is some sort of magazine audience or "rating" survey.

With some additional coaching about how to use the board, you go through the cards, sorting them into the appropriate answer boxes. Finally you finish. At this point, there are 4 cards in the *sure read* pile and another 2 in the *might have read* pile. Although you aren't really certain about every one of your sortings, the interviewer isn't pressing you to reconsider your answers. Rather, another deck appears (this time with light blue cards). It is shuffled and then you sort through 40 additional titles, only to find that after this, yet another deck is produced (this time the cards are gray).

At last the card sorting exercise ends. You've reviewed 120 titles, most of which were classified in the *sure not read* category. However, 15 cards are located either on the *sure* or *might have read* boxes.

Continued

Magazine Audience Measurement Continued

Next, the interviewer removes all of the *not read* cards. As you watch, the remaining ones are sorted, with weeklies separated from bi-weeklies or monthlies. Now these titles are read one by one as they appear on the answer board, beginning with the weeklies. As the first weekly's card comes up, the interviewer asks a sequence of questions. These begin with how many (if any) out of four issues of this magazine you read or looked into. Then there is another "audience" question. This time the interviewer wants to know whether you read or looked into any copy of this magazine in the past seven days.

You do your best to answer, trying to imagine all of the recent issues you may have seen and, following this, whether any were read in the past week. In the first sequence, you report reading two of four issues—a guess since you don't subscribe to or regularly receive the magazine. In the second sequence you state that you aren't sure whether you read a copy in the past seven days. The interviewer notes this and moves on to the next weekly. This time it's a magazine you *do* subscribe to and enjoy reading. Surer of your answers, you report seeing all of the last four issues though, for the life of you, you can't recall what was on the covers of half of them, let alone their contents.

Fortunately, the interviewer doesn't ask about this. Rather, the questions turn again to the past week time frame; in this case you happen to claim readership. Now, however, the interviewer continues with questions you weren't asked for the first magazine. Why? You are not really sure. The reason, of course, is that this is the first time you qualified as an average issue reader, but this distinction isn't explained. Instead, you are asked to estimate how much time you spent with the last issue you read, how many different days you were exposed to it, where you saw it, whether it is one of your favorite magazines (or a choice of other less positive ratings) and any reader actions you recall taking (coupon clipping, tearing out ads, writing to the editors, etc.).

You answer to the best of your ability and the interviewer seems satisfied. Then, after a similar interview for another weekly, you move on to the next pile of cards, which are monthlies. The first card triggers the usual questions. Again, you are asked how many of four issues you read or looked into.

The magazine is one you really like and see quite often. In fact, you bought the last issue at a newsstand and have considered subscribing to the publication. At any rate, you must have looked at this particular copy three or four times since you bought it. But as for older issues? Hmm? How many out of four do you read? You're not really sure.

The interviewer leans forward, looking at you fixedly. You answer: three issues. Yes, I read three out of four issues. This is noted. Then the interviewer asks whether you read or looked into any issue of this magazine in the past 30 days.

Continued

Magazine Audience Measurement Continued

Why 30 days, you ask yourself? For the previous magazines they asked about past 7 days readership.

Gently but firmly, the interviewer encourages you to move on. Yes, you reply. I've read it in the past month. Then there are more queries about reading time, where you read the magazine, etc.

The next publication—also a monthly—is not one of your favorites. Indeed, as the interviewing process begins, you wish you hadn't placed it into the *might have seen* pile in the initial card sort. But now it's too late, so you estimate the number of four issues you see as one. Then you change it to two. In both cases, your intent was to signify *not often*. Now about the past 30 days. They're talking about any issue here. Well, maybe...after all, I just said I read two out of four issues...OK, make it a "yes."

And so the interview continues. By the time it concludes, 50 minutes have elapsed and you have provided audience determining replies for 15 magazines. In some cases you were almost positive about your answers, but in others you didn't feel confident at all. In any event, you hope your participation in this project proves valuable to those conducting the survey. And, by the looks of her, the interviewer is certainly pleased as she packs up and prepares to leave.

These hypothetical scenarios provide a glaring enough contrast between the burdens and complexities imposed on respondents by through-the-book and recent reading, but consider what these would be like if the through-the-book interview was extended from 60 to 120 titles, while the recent reading interview sought to measure 250, not 120 magazines. Imagine the consequences for heavy magazine readers—one-fifth of the sample that accounts for 40-50% of all readings. In the case of an expanded through-the-book study covering 120 publications, heavy readers could screen in 20 titles; for an expanded recent reading study (250 publications) the heavy reading respondent might screen in 40+ titles. In both circumstances this would place a tremendous burden on both participants—the interviewer and the respondent—and one or both might try to speed up the process to the detriment of the study.

Critics of through-the-book have always maintained that this methodology understates audiences, particularly of monthlies. This is due to the inherent memory loss that occurs when people are shown a copy of a weekly that is 5 weeks old or a monthly that is 10 weeks old. The manner in which magazines are "stripped-down" can also contribute to this problem. A fair representation of an issue's editorial contents may not be attained by showing only 10 or 12 items; moreover, it may take more items to adequately portray the editorial content of a thick "reading" magazine relative to a thinner, "fast read" pictorial book. Finally, there is the likelihood that additional first time readers are attained well beyond the arbitrary 5- and 10-week age limits that through-the-book uses for weeklies and monthlies. Scattered throughout the literature and the findings of various studies, are bits of evidence that support this

Continued

Magazine Audience Measurement Continued

contention. For example, in the old Politz reading day studies conducted in the mid- to late-1950s, day by day through-the-book surveys determining “yesterday” readership suggested that significant reader attainment might occur as late as the 10th and 11th week after a monthly’s publication. More recently, an Audits & Surveys experimental mail study using the frequency of reading method observed 20% higher reader-per-copy levels when the pictorial cover representation used in its screening process was for 16-week-old issues rather than those aged 8-10 weeks.

The arguments against recent reading can be equally vehement and telling. Defenders of through-the-book point out that their method disguises its true intent to the respondent by posing as an editorial interest not an audience study, while this is not the case for recent reading. Moreover, by showing respondents an actual copy, even in skeletonized form, and then inquiring specifically about it, through-the-book produces a much sharper memory test than recent reading, which relies excessively on a respondent’s recall prowess, without the benefit of a visual reminder.

As for the differences in readership claims for weeklies and monthlies, critics of recent reading believe that this method causes respondents to confuse repeat readings with first time exposures when reconstructing the totality of their past week or past month readership in a hurried interview involving numerous titles. Since monthlies generate more repeat exposures, this favors monthlies unduly over weeklies. In addition, there is the issue of interviewer participation. In through-the-book, there is an interviewer-controlled screening process (a booklet with three logos per page is flipped page by page by the interviewer for the respondent’s perusal). As each magazine screens in or out, the interviewer asks the respondent if he or she is sure about past six months exposure, in the process permitting, or even encouraging a correction in the claim if this seems indicated. In recent reading, however, the respondent plays a card sort game, using a scoring board. With so many cards involved, it is questionable that the respondent will go back over the full *sure* and *might have read* piles to weed out overclaiming.

Once the “audience” questions commence, the differences between through-the-book and recent reading are even greater. The former, rather incidentally, asks respondents who have screened in and examined a skeletonized magazine whether they have read or looked into a copy of that issue before. With recent reading, in order to produce multiple issue reach and frequency estimates, respondents are first asked to approximate how many of four issues of the magazine they read or looked into—a generalized frame of reference that extends well beyond the average publication’s issue cycle. Someone may estimate that s/he typically sees two of four issues of a monthly, including exposures that may have taken place as long as three or four months ago, as well as more recent contacts with the magazine. What happens next,

Continued

Magazine Audience Measurement Continued

when the actual recent reading question is posed? Does the respondent who has already claimed *sure* or *possible* exposure in the past six months and then answered yes to the frequency question (by reporting two issues out of four are read) feel unduly obligated to “confirm” his/her prior replies with a “yes” on past month readership, even when this, on careful reflection, may not be the case? Is it reasonable to expect that the interviewer, knowing that many other magazines must be asked about, will encourage such corrective ponderings whenever they arise?

The most compelling evidence of the inflationary aspects inherent in recent reading is offered by the varying results the methodology itself has produced. Initially, in the TGI studies, the recent reading queries were deeply placed in a self-administered questionnaire. Logos were not employed to screen in respondents and the magazines were simply listed, with weeklies, which came first, separated from monthlies. Even so, TGI’s audience projections were 15-25% higher than through-the-book. When MRI introduced its major refinement, with an interview supplanting the self-administered questionnaire, and the introduction of the card sort screening game, another 20-30% increment in readership claims ensued, with some of the mid-sized and smaller monthlies “gaining” almost twice as much. Then, in 1995 when SMRB switched from through-the-book to a “streamlined” version of recent reading (where interviewers went promptly to the magazine questions with a minimum of delay), audiences rose another 10-15%. Finally, Nielsen’s Home*Scan, with its self-interviewed respondents using electronic scanners to record their answers, witnessed another quantum leap, producing levels 25-30% higher than SMRB’s streamlined recent reading interviews and 40-45% higher than MRI’s findings.

Clearly, the recent reading questioning sequence and the nature of the answers that it generates, is extremely sensitive to varying executional elements, especially the respondent-interviewer interfacing. Moreover, the core assumption in recent reading, that any issue readership in the past week for weeklies and past month for monthlies is the equivalent of average issue audience, has never been put to the test. This also applies to the assumption that respondents can provide equally accurate recall answers regardless of what time frame is inquired about.

To sum up, SMRB’s execution of through-the-book pushed this method well beyond its natural limits and cannot be set up as an “ideal” basis of comparison, but recent reading hardly qualifies as a more precise or superior methodology. It has evolved in the U.S., if not elsewhere, primarily as a direct competitor to through-the-book, gradually assuming the latter’s trappings (interviews replacing self-administered questionnaires, logo card screenings, etc.) while seeking to retain its original advantages (the ability to measure many more

Continued

Magazine Audience Measurement Continued

magazines and obtain more data about them at a lower cost). The resulting hybrid, while incorporating some of the better points of both methods, has created an overburdened recent reading interviewing system which probably leads to over-claiming for monthlies relative to weeklies, beyond what a “perfect” study might find.

What do magazine audience studies measure? Frankly, no one is really sure. To the extent that respondents understand the questions and are able to comply, the surveys generate much of the information their designers intended to obtain. A good part—though by no means the full part—of the answers reflect true readership. However, their precision certainly is questionable. The respondent may be thinking about several recent issues, or older editions, while repeat readings, particularly of monthlies, may further cloud the picture. Add to this the all-encompassing broadness of the readership definition, plus variables such as title confusion, attempts to impress or fool the interviewer (by “closet readers”), failure to correct misstatements due to pressure to complete the interview, compounded by errors caused by multiple forms of audience questioning, memory loss, various forms of scaling bias (some types of respondents are inherently more precise in their answers than others), etc., and it is obvious that the audience surveys provide only the crudest of benchmarks. Moreover, their biases and errors, whatever they may be, do not necessarily apply equally to all of the publications involved. Used with caution as directionally indicative of audience size (high or low reader-per-copy velocity), demographic targeting capabilities or reach/frequency attainment, these studies are helpful. But taken literally as absolute “audience counts,” producing definitive and minute distinctions between publications (relative to ad page costs), they can lead to bad, or more properly, to unfair decisions by space buyers. In short, as currently constituted syndicated audience studies ask too many difficult-to-answer questions from too few respondents to produce totally reliable readership findings. ■