THE BENEFITS OF A STANDARDIZED SYSTEMS APPROACH TO COPY RESEARCH

by

Joel S. Dubow
Communications Research Manager
Coca-Cola USA

Third Annual Advertising Research Foundation Copy Research Workshop
The New York Hilton
May 28-29, 1986

When Joe Eastlack asked me to speak at today's workshop on "A Standardized Systems Approach to Copy Research," I readily accepted the invitation. After all, I had exactly that speech in my files under a different title. Joe told me he was familiar with that speech, and reminded me that I had delivered it at this very same workshop two years earlier. He went on to note that in my earlier speech I described how we test our advertising, and why we do so in the way that we do, and how we validated the system. But he also commented that I was a bit circumspect in the details of our validation study; and, that I had not spoken at all about how our testing system impacted our advertising decision-making process. Joe indicated that perhaps now, two years later, some of the data in our files had become sufficiently outdated that I could be a bit more open. We decided that I would interpret the title he had given to me to read as "(The Benefits of) a Standardized Systems Approach to Copy Research."

If all of you had attended the 1984 workshop, I could launch right into those benefits. But for the benefit of those you who did not attend in 1984, I'll begin with a review of our system.

The major copy testing system which we use at Coca-Cola, and which has been in operation since the mid-1970's, was developed to assess image advertising. It addresses the two common elements so frequently discussed among advertising researchers: recall and persuasion. But it interprets recall in an atypical manner, and assesses persuasion via a unique testing method. In fact, we consider those differences to be of sufficient magnitude, and in the context of image advertising to reflect a sufficiently different philosophy of how advertising works, that we have come to rename them.

We refer to recall as "intrusiveness," and we measure intrusiveness because we believe, in order to be effective, the advertising must "intrude" into the target audience's consciousness. In the context of image advertising we refer to persuasion as "communications value," and we measure communications value because we believe that once the advertising intrudes, it must communicate about the brand what it is we mean to communicate.

We measure intrusiveness using one of the popular, standardized day-after-recall tests: Gallup and Robinson's In-View service. But when it is image advertising that is at issue we interpret the results a bit differently from the usual copypoint playback approach. Where image advertising is at issue, we have found that copypoint playback is virtually useless -- it begs the obvious, while understating the subtle. (It may be this understating of the subtle that causes much of image advertising to score below recall norms for copypoint playback -- consumers cannot easily put the subtle into words.)

What is important to us in the intrusiveness test is that we find evidence that the content of the commercial has become associated with the brand. The evidence comes from the responses to the questions:

Do you recall seeing a commercial for Brand X?
What did they say in the commercial?
What did they show in the commercial?

The verbatim protocols which occur in response to these questions are examined for proof that an association has been formed between our brand name and the content of the commercial; or, if you will, that a "cognitively significant intrusion" has occurred. (Please note that I use the word "intrusion" in a non-pejorative sense; image advertising is rarely unwelcome.)

Finding a measure of communications value was not nearly so easy as finding a measure of intrusiveness. The existing systems all relied on a single measure of "persuasion," usually a pre-post change in buying intentions. We needed a system which would address the many facets of imagery called for by our brand strategies. And, furthermore, we doubted that an overly rationalizable buying decision question would be a valid measure of the effectiveness of image advertising.

It did not take us long to realize that we would have to develop our own technique for measuring the communications value of commercials. We had a fairly good idea of the form it would take, but we needed some expertise on how to make it work in the field. For that we called on Kenneth Hollander Associates, an Atlanta-based research supplier familiar with the mind sets that existed at both Coca-Cola USA and McCann-Erickson. Ken had a few years earlier been at McCann-Erickson and worked on the Coca-Cola account.

What emerged from these Coca-Cola/McCann-Erickson/Ken Hollander discussions came to be known as our "Quantitative Communications Test," or more simply "QCT." Let me describe it to you in conceptually simple terms: The QCT uses a test versus control design. The Control Group is recruited in malls to age/sex/brand usage criteria that fit our target audience. Once selected all they do is rate the subject brand on 30-40 brand image items. The Test Group is recruited to the same target specifications, shown the commercial (twice), and asked to rate the commercial on several brand attributes such as believability, interest value, etc. Then they are also asked to rate the brand itself -- on the same set of brand image items on which the Control Group rated the brand. Any difference between the brand image ratings of the Test Group and Control Group is taken as the communications value of the commercial.

Depending on the specific brand we categorize the image items into "intrinsic" items, "extrinsic" items, and "category items." "Intrinsics" refer to the things that comprise soft drinks and what soft drinks do for you. "Extrinsics" are the subtleties that copypoint playback rarely registers -- the brand personality and its user image. "Category Items" refer to special items relevant to the specific type of soft drink being tested: low calorie, or cola, or lemon-lime, or juice-added, etc.

By including up to 40 items in the QCT we obtain very good diagnostic information about just what any given commercial does and does not communicate. And, we use large enough sample sizes to determine to whom it communicates -- males versus females, or young versus old, or our own users versus competitors' users -- and what it communicates to them.

Ultimately we take about a dozen of what we consider the most important strategy items and amalgamate them into a Total QCT score, which we then adjust based on the G&R intrusiveness score, to give us an "Effective" QCT

Score. Over 10 years we have amassed a data base of scores for nearly 800 of our own and our competitors commercials. We test virtually every commercial we produce and about three out of every four competitive commercials that we manage to lay our hands on.

The benefits of having a standardized system for copy testing begin to emerge as you develop a data base like ours. For us that occurred after about three years when we had enough of a history to conduct a validation study.

I will begin with the second benefit we derived. When I addressed this workshop in 1984, I described the results of our validation study something like this:

> We developed the ability at Coca-Cola USA, under certain marketing conditions, to rather accurately predict market share for some of our brands as a function of advertising weight and promotional activities. We performed that analysis for a three year period, applying our data to 18 consecutive Nielsen bi-monthly tracking periods. In some periods we enjoyed a higher share than our model predicted, while in others we suffered a lower share than anticipated. When we compared those deviations to the scores for the concurrent copy, we found a definite tendency for positive share deviations to be associated with above average copy scores, and for negative share deviations to be associated with below average copy scores. If you call a significant positive deviation a win, a significant negative deviation a loss, and an insignificant deviation from expected a tie, consider these W-L-T records:

	Win	Loss	Tie
With Above Average Copy	4	0	2
With Average Copy	2	1	1
With Below Average Copy	1	6	1

But the actual benefit we derived from the validation analysis went beyond satisfying our management that the system worked. It came directly from the multiple regression formula that emerged from the stepwise linear regression analysis:

$$\text{Brand Share} = a\,\text{Share Advg} + b\,\text{Share Promo} + c\,\text{Copy Scores} + K$$

The coefficients a, b, and c literally told us the worth in incremental share, and incremental sales, of unit increases in advertising dollars, promotional dollars, and copy test scores. That generated a cost effectiveness analysis. We were able to directly calculate the costs of gaining share either through increases in advertising or promotional spending. We were also able to impute the cost of improving our average copy test scores for aired commercials by simply overproducing commercials and throwing the weaker half away. The net result, at that time, was that given a hypothetical $5,000,000 to spend, the best thing we could do with it was put it all into advertising production. In fact, the next year our management did commit to substantial increases in our production budgets.

The first benefit we derived was one we were enjoying all along, but could not verify until we saw the regression coefficient for copy scores. At that time we tended to use pools of three commercials. Historically we would rotate them one-third each, but once we set our copy testing system into place we altered the rotation to one-half, one-third, and one-sixth, for the strongest, middle and weakest commercials. We calculated the net effect on average copy scores as actually aired for the three year period under study, and placed that increment against the regression coefficient for copy scores. We calculated the sales benefit, and extracted the margin on those incremental sales. As it occurred, the incremental profit earned merely by rotating commercials according to copy test scores was just about the cost of operating the system during that three year period. On that basis alone, the copy testing system paid for itself -- all other benefits were gravy.

A third benefit of the regression analysis that emerged from the validation study is a bit arcane. It occurred when we separated out the QCT and G&R components of our Effective QCT Score. We found a linear relationship between QCT (read: persuasion) and sales impact, but a non-linear relationship between G&R (read: recall) and sales impact. I recently had the occasion to review the literature on recall testing, most of it critical of the technique, and I realized that most of that criticism was based on the assumption of a linear function. The implications regarding the utility of recall testing are quite different if the underlying function is markedly non-linear. It will be a few more years before our actual data on the shape of the recall function becomes sufficiently dated to share it with you. I'll leave you wondering about it, so that perhaps in another several years I'll be invited back to address this conference again.

A fourth benefit we derive from standardized and continual testing, especially because we test competitive commercials nearly as thoroughly as we test our own, is that we usually know before our competitors do, how changes in their advertising copy will impact their market share.

I offer two examples. The first occurred in 1978 when Dr Pepper switched from their "World's Most Original Soft Drink" campaign to the famous, highly touted "Be a Pepper" campaign. Our testing suggested that "Be a Pepper" matched the high intrusiveness levels of its predecessor, but scored terribly low on QCT. So low that we believe Dr Pepper in effect suffered more than a 50% drop in advertising effectiveness. Not long thereafter, Dr Pepper's steady five year long pattern of market share growth flattened out and then began to erode.

A second example comes from 7-Up. Prior to 1979, 7-Up was running a quite representative campaign called "Undo It." In 1979 they changed to a wonderful campaign called "America is Turning 7-Up," with G&R scores only 60% the strength of "Undo It," and QCT scores only 60% of the strength of "Undo It" -- a net drop in Effective QCT Scores of nearly two-thirds. They also increased 7-Up's annual spending by 70% in 1979. In order to overcome the copy deficit they created -- and did not know about -- they would have had to have tripled it. The record shows that, with the onset of "Turning 7-Up," that brand began a steady three year share decline.

We've also benefitted from our testing system by avoiding a few shortfalls of our own. Consider this never-aired Coke campaign that was intended for early 1979: "Things Go Better When You and Coke Get Together." When that campaign fell short of our standards, we reverted to a back-up, six month extension of "Coke Adds Life," based on other copy testing results. We had recently tested a number of our international commercials, amended to utilize domestic music and lyrics. Several proved promising, and with some minor editing to make them useable on American television we extended the Coke Adds Life campaign through mid-1979.

One of those international commercials was particularly inspiring: "Smiles." It scored extremely well and served as the inspiration for the campaign we launched in late summer, 1979: "Have a Coke and a Smile." And I might add, that had we not known [nay, could not have known without our testing system] that "Things Go Better When You and Coke Get Together" fell below our standards, and that "Smiles" was so strong, we would never have travelled the path that soon brought us all this commercial: "Mean Joe Greene."

The benefits of a standardized copy testing system derive from the history it provides, and the opportunity to compare what is new to what has happened before. To ignore history is to repeat its mistakes. If you don't believe that just ask any brand manager.