Continuous Likability Measurement

Use this potent technique for developing effective TV ads.

By William H. Murphy and Sidney S.L. Tang

According to 1980s’ figures, expenditures for TV ads account for about one-fourth of the $100 billion plus advertising industry. The current environment requires increased accountability of television ad resources—of late, companies and advertising agencies are under immense pressure to create effective ads. Companies must conduct research appraisals to gain confidence, and to protect themselves from potentially negative, public reactions to their ads.

Most ad agencies use some form of preliminary research in ad campaign preparation—the majority evaluate copy ideas or use storyboards, and many evaluate rough commercial takes. Almost all agencies use appraisals to evaluate finished commercials.

Traditionally, finished commercial appraisals involved post-viewing measures, including recall, recognition, and likability. Yet, concerns with post-viewing measurement do exist. First, these methods require viewers to reflect on their experience. This introduces uncontrollable elements into the mix—leading to measurement validity questions. Second, post-viewing measurement is suited primarily for “go/no go” decisions, despite most practitioners’ view that ad testing should provide creative guidance, within an entire body of research.

It’s the variations’ effectiveness in the temporal occurrence of execution elements that should be key in ad testing; this permits advertisers to determine the particular moments in an ad that are effective with target audiences, and to make presentation changes as a result.

Online continuous likability measurement copy testing can be used to evaluate specific ad moments, enabling advertisers to gauge the effectiveness of presentation variations. Also, different presentation variation effects can be assessed across viewing audience segments. Thus, online measurement offers far more information than the “go/no go” decision; it helps decision makers with creative guidance, and ad design.

Continuous Likability Measurement

Liking (i.e., a positive attitude) is important to advertisers because peripheral route processing can
be “more true” in the consumer’s mind. To paraphrase Ian Fenwick and Marshall D. Rice, from their 1991 *Journal of Advertising Research* article: By concentrating on creating affective advertising executions, firms can “inoculate” a brand against competitors with superior product attributes. The idea: Consumers form a positive attitude toward the brand because they like the advertising, making them impervious to counter advertising that emphasizes attribute superiority.

And likability does seem to perform better than recall or persuasion. Inferences about liked ads, in post-viewing copy testing, include: (1) liked ads are noticed and remembered, (2) liked ads are associated with positive brand attitudes, and (3) liked ads are associated with ad persuasiveness. Online likability measurement offers insight into fleeting emotional responses that occur during message viewing—especially useful, given that emotions form and change quickly. Emotional changes that occur during exposure are identified through continuous measurement, focusing attention on effective vs. less effective scenes. Advertisers gain insight into where, and how, to fine-tune their ad messages. Measurement tools that rely on post-viewing evaluations don’t provide this “real time” sense of the effect of ads on emotions. Post-exposure copy testing is overly rational, requiring excessive verbal interpretation from viewers.

Online continuous likability’s specific strengths include:

- Microdiagnosis capability where continuous measurements, taken in “real time,” provide information into an ad’s effectiveness.
- Copy segment evaluation where it’s possible to identify and evaluate reactions to specific stimuli introduced in various segments of a commercial.
- Less attenuation in measurement validity because continuous measurement doesn’t require a subject to think about and crystallize beliefs (this is required in recall and other post-exposure measurement approaches).
- Good test-retest reliability.

**Current Practices**

A number of systems for continuous likability measurement copy testing are available. A 1991 advocacy article focused on the program evaluation analysis computer (PEAC) system. In it the authors describe the PEAC method: “...Subjects are asked to press buttons on a hand-held unit to indicate the degree to which they “feel positive” about what is on the television screen at that moment. The unit has five buttons lettered A through E located in a vertical column. Subjects are told to use the buttons as often as they wish according to the following scale: A = very positive, B = positive, C = neutral (defined as no strong feelings), D = negative, and E = very negative. Subjects are given explicit instructions during this procedure not to rate the product in the commercial but the commercial itself.”

For the copy test, respondents are selected on the basis of client specifications. During exposure, the hand units record responses every two seconds—the responses are sent to the computer. After tabulation, a graphical trace is displayed (see Exhibit 1 for a typical trace).

Generally, analysis information comes from two sources: the trace and participant interviews in post-measurement sessions. Advertisers visually examine the trace, looking for meaning from fluctuations. They review participants’ comments from the post-measurement session, enhancing their understanding of trace changes. Comments are frequently obtained during a post-exposure/post-data-collection focus group session where the trace and the ad are shown, simultaneously. During the session, subjects are asked to think about changes in the trace and why their pattern of response changed at certain moments (time points) in the ad.

Unfortunately, the satisfaction of the continuous measurement copy testing method ends here today. Its procedures and accompanying interpretations
help in understanding the overall effects of advertising stimuli on an audience; however, current practice gears analysis at a macrolevel similar to other copy testing methods, emphasizing the “go/no go” method. Further, analyses are post hoc; they don’t purposefully interpret the effects of advertising stimuli. Unless the trace calls attention to it, the purposeful scene changes and their effects on an audience are ignored. Even when the trace calls attention to a change in emotional response, inference is taken from comments at the focus group sessions following measurement. (We find that these group respondent interviews often reveal little more than the opinions of “discussion leaders” among respondents.)

Despite having purposefully designed ad scenes (and scene changes), and continuous likability measurement data, lacking are significance tests of changes in response to critical scenes. Yet, these data are available to conduct analyses.

Making It Better

Continuous likability measurement becomes more purposeful if copy testing depends on a priori expectations, not post hoc interpretations.

Completing a full-fledged TV ad is a purposeful activity, providing audience exposure to environments, scenes, linguistic nuances, and so forth. Since ads aren’t developed randomly, why should their effects be analyzed randomly? Analyses of “critical moments” in ads should be an integral part of copy testing. Advertisers should make specific diagnoses that will enhance the overall performance of a commercial—and provide their clients with supportive evidence regarding ad decisions as well.

Advertisers should anticipate and analyze differences in various target audience responses to critical ad moments. By doing so, advertisers can refine and enhance the effectiveness of marketing messages for intended audiences. Here again, advertisers could substantiate an ad’s effects for client reporting purposes.

During the copy development process, clients expect explanations of why specific characters, music, and scenes are used. The next step is to lay out specific expectations, in a scene-by-scene fashion, anticipating effects on the target audience. (We recommend these expectations be clarified before conducting continuous measurement tests.) Exhibit 2 shows a typical commercial profile (prepared prior to continuous measurement tests) and includes the timing for key scenes in the ad and the expectations for each scene.

After completing the commercial profile, and specifying audiences to target, subjects are selected for continuous copy testing measurement appraisals. Then, statistical analysis is used to evaluate (quantitatively and objectively) the design’s effectiveness. Statistical analysis provides empirical evidence as to why specific scenes should be kept, dropped, or repositioned for the final copy.

Gains from this copy testing strategy are numerous:

- Advertisers can examine anticipated, specific effects from “critical moments,” usually scene changes, more closely.
- Copy testing no longer needs to rely on post hoc analysis since a priori expectations are stated.
- Statistical analysis can provide an objective empirical insight into the research data.
- Explanations to clients, before and after copy testing, can be more precise and objective.
- Information gains are evident since answers to a priori questions provide the basis for discourse with clients.

**INFORMATION GAINS**

As suggested, though continuous measurement is underutilized, it can potentially provide...
far more powerful information to advertisers. To realize this potential there’s a statistical method that enables researchers to discern continuous measurement data quantitatively.

The Statistical Method

The purpose of analysis, and choosing a statistical method, is to interpret the meaning of the collected data. In continuous measurement, the collected data are measurements taken during audience exposure to an ad; a datum represents each subject’s response at a time interval. Depending on the commercial’s length and the response measurement’s frequency, the number of measurements repeated throughout the commercial varies. For example, in a 30-second commercial, data collection every other second yields 15 repeated measures for an individual respondent. In analyzing a particular segment of a commercial, a subset of measures corresponding to the segment are treated as dependent variables.

Categorical data analysis is chosen for the interval measurement scale used with the PEAC method (i.e., the degree to which subjects “feel positive”). Because of the scale’s discrete nature, it’s not appropriate to directly apply other general linear models designed for continuous data. (Importantly, statistical inferences made on the basis of an erroneous model are likely to be invalid and misleading.) Categorical data analysis offers an array of analytical techniques designed for discrete data.

Categorical data analysis doesn’t require normal distribution; it requires a product multinomial distribution assumption. Each distribution “category” has its own population probability, and all categories’ probabilities sum to 1. In mathematical symbols, the index nomenclature for the product multinomial distribution is: \( \text{Prob}(n_1, n_2, ..., n_r) = (n_1 + n_2 + ... + n_r)! \left( \frac{p_1}{1} \right)^{n_1} \left( \frac{p_2}{2} \right)^{n_2} ... \left( \frac{p_r}{r} \right)^{n_r} / n_1! ... n_r! \) where \( n_r \) is the number of occurrences of category \( k \); \( p_k \) is the probability of occurrence for category \( k \); \( p_c = 1 \); and \( n = \text{sample size} \). With a large enough sample size, the distribution is asymptotically normal, thus many significance testing techniques can be applied in categorical data analysis.

In categorical data analysis, data are transformed through a response function \( F(p) = X\beta \), where \( p \) is a probability vector, \( X \) is a design matrix with constants indicating the model’s structure, and \( \beta \) is a vector of to-be-estimated parameters. The response function can be defined many ways. Popular definitions include logit \( \ln\frac{p_j}{1-p_j} \), cumulative logit \( \ln\left[ 1-p_j/p_{jl} \right] \), and simple average (for data with at least an ordinal scale).

When a model is fitted to the data, transformed data variations are divided into the model space (i.e., variations that can be accounted for by the model) and the error space (i.e., variations that can’t be accounted for by the model). Wald’s chi-square statistic has a chi-square distribution when the null hypothesis is true; it’s used to test the goodness-of-fit of the model, and the significance of main effects and interaction effects, if any.

To control variations between subjects, repeated measures method is applied (its design is a simple modification of the model equation described above). In a repeated measure setting, the design matrix \( X \) and the vector \( \beta \) reflect the presence of multiple dependent variables (i.e., successive continuous measurements). The effect of repeat measurement is incorporated into the model being tested. Thus, the model contains effects from the main factor(s), repeated measurement, interaction(s), if any, and the residual. By dissecting the data into model and error space, variation sources are studied and the most appropriate data model determined.

The technique requires that the underlying distributional assumption of the categorical data models be asymptotically normal—if the sample size is large enough, 30 or above for each response function, the data will conform to a normal distribution. The number of response functions for generalized logit is \( r-1 \), since the probability of all categories add to 1. For example, if the measurement scale is nominal and contains three possible categories, \( r = 3 \), (e.g., positive, neutral, and negative), then the sample size for each population (e.g., male and female) should be \( 30 \times (3-1) = 60 \) or above. If an “average” response function is used instead of generalized logits (e.g., population means of the positive, neutral, and negative categories), the sample size for each population (e.g., male and female) should be 30 or above.

**THE REPEATED MEASURES METHOD**

The repeated measures method is a powerful tool for analyzing continuous likability measurement data. It’s useful for investigating emotional response differences, to specific stimuli, between subgroups of a targeted audience. For instance, this method can discern particular stimuli response differences between males and females in a target audience. (The client, through the advertising agency, selects appropriate respondents for the
copy test.) Then, with data collected, the trace is revealed (see Exhibit 3).

The impact of a particular scene change, where the effect of the scene on males vs. females is expected to show a priori at time points 9-12, can now be examined statistically. This analysis can be conducted at any “critical moment” where the client anticipates a difference. Instead of relying on gut feelings to measure audience responses (e.g., Does the male response seem to differ from the female response?), researchers can use a statistical tool that indicates whether the difference is statistically significant. Thus, continuous measurement data can be analyzed for all kinds of target audience breakdowns—gender, age, income, and so forth—that the client deems relevant.

This application helps advertisers refine marketing communication messages by analyzing subgroup responses to specific time periods of a commercial. Data collected at preselected time intervals within a commercial are treated as the dependent variables (repeated measures). The effects of subgroups and repeated measurement of dependent variables are studied in the “best-fit” model development process.

A second application involves executional effectiveness differences (on the basis of emotional response) of different executions of a segment in a commercial copy. For example, a client wants to compare the effects of a male vs. a female voice-over. As before, the client receives a trace of each execution, perhaps overlain as in Exhibit 4; however, now a quantitative method helps to complement the decision makers’ intuition in examining the differences.

If the client chooses, analysis can also be conducted using the first application’s procedure. Thus, analyzing response difference by target audience subgroups, according to structural change in execution, is also appropriate here.

The method is ideal for quantitatively evaluating and comparing different copy executions during the commercial’s development stage. For example, an advertiser can determine whether different presentations of a key stimulus are better received by particular target audience subgroups, and whether statistically significant differences exist between subgroups’ responses.

Data collected in a particular segment of a commercial, and in a corresponding commercial, are treated as dependent variables (repeated measures). Group and repeated measure effects can then be investigated.

A third application helps to determine whether a particular scene is more effective in one position vs. another position in the ad. It can be used to study the effects of scene ordering in a highly visual automobile ad, for example. Here, the client wants to examine the effects of placing a particular scene early in the ad vs. late in the ad. The two traces are shown together in Exhibit 5.

**A Hands-On Example**

We’ve introduced a number of repeated measures method uses. The method only comes to life, however, in practice where we refer to the analysis as critical moment analysis (CMA). The example follows (we don’t use the client’s name due to confidentiality requirements).
The primary objective behind the research: To identify differences between target markets in reacting to key messages within the test commercial. The research supplier proposes several research options—among them, the continuous measurement method. Its advantages, especially the real-time measure of respondents' feelings, leads to the "go-ahead" to use it for the research project.

**Research process:** The research firm collects data from two respondent groups: males and females. Respondents are recruited according to the client firm's specifications. Using the PEAC system, respondents are given hand-held units to continuously record their feelings towards the commercials as they view them in a theater setting. The responses, ranging from "very positive" to "very negative" in a 5-point scale, are recorded in 2-second intervals and registered in real time by a computer system. Exhibit 6 contains the session's results.

**Data analysis:** The research objective includes analysis of the continuous measurements in certain segments of the commercial. According to the client, these segments, or "critical moments," contain key messages. To study group differences for these segments, the research team proposes the CMA method.

One of the commercial's key messages occurs in time interval 9-12. The goal is to statistically determine whether significant differences exist between male and female reactions to the message during this interval. The research team applies a statistical method, categorical data analysis, to the data. Specifically, the model $F(p) = X\beta$ is used, where $F$ is a response function, $p$ is the vector of population proportions, $X$ is the design matrix for the model, and $\beta$ is the vector of population parameters.

The best-fit model is chosen using the residual criterion. A variance analysis reveals that the residual's p-value is 0.19 (see Exhibit 7). And the model reveals statistically significant differences between gender groups ($p$-value = 0.01). The result provides researchers with diagnostic information about how the targeted audience subgroups react to the key segment's stimulus/message. Similar analyses are conducted for each of the commercial's key messages. On the basis of the analyses' findings, the client gains insights into subgroup differences (gender differences) in reaction to the key messages, enabling the client to evaluate the commercial with greater precision.

**THE FINAL ANALYSIS**

As indicated, continuous likability measurement captures unique information about a commercial
that's not available from post-viewing appraisals. Unfortunately, most analyses use continuous likability in much the same way as post-viewing measurement—as an appraisal for the “go/no go” decision.

The measurement methodology discussed here offers substantive insights into (1) whether differences exist in reactions to commercials by target audience subgroups; (2) whether differences exist in audience response to various executional changes within a segment of a commercial; and (3) whether changes in the positioning of key scenes within the commercial are associated with significant differences in target audience reaction.

Use of continuous likability measurement encourages use of a systematic copy research plan. And CMA, as part of a systematic research approach to copy testing, provides a quantitative and objective way to examine a commercial.

Overall, this strategy can lead to improved ads—ads that help sell the product, ads that help the audience remember the product, and ads that improve the audience’s product recall by improving their ad recall. It’s our impression that continuous measurement copy testing will see increased use in years to come, as a viable tool for the measurement of emotional response to advertising.

### Exhibit 6

**Continuous copy testing results**

<table>
<thead>
<tr>
<th>Time point (2-second interval)</th>
<th>Aggregate sample (n=73)</th>
<th>Female (n=38)</th>
<th>Male (n=35)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.15</td>
<td>3.16</td>
<td>3.14</td>
</tr>
<tr>
<td>2</td>
<td>3.15</td>
<td>3.24</td>
<td>3.06</td>
</tr>
<tr>
<td>3</td>
<td>3.38</td>
<td>3.47</td>
<td>3.29</td>
</tr>
<tr>
<td>4</td>
<td>3.60</td>
<td>3.79</td>
<td>3.40</td>
</tr>
<tr>
<td>5</td>
<td>3.59</td>
<td>3.79</td>
<td>3.37</td>
</tr>
<tr>
<td>6</td>
<td>3.63</td>
<td>3.89</td>
<td>3.34</td>
</tr>
<tr>
<td>7</td>
<td>3.55</td>
<td>3.79</td>
<td>3.29</td>
</tr>
<tr>
<td>8</td>
<td>3.66</td>
<td>3.84</td>
<td>3.46</td>
</tr>
<tr>
<td>9</td>
<td>3.88</td>
<td>4.06</td>
<td>3.66</td>
</tr>
<tr>
<td>10</td>
<td>3.84</td>
<td>4.18</td>
<td>3.46</td>
</tr>
<tr>
<td>11</td>
<td>3.84</td>
<td>4.13</td>
<td>3.51</td>
</tr>
<tr>
<td>12</td>
<td>3.84</td>
<td>4.18</td>
<td>3.46</td>
</tr>
<tr>
<td>13</td>
<td>3.77</td>
<td>4.13</td>
<td>3.37</td>
</tr>
<tr>
<td>14</td>
<td>3.58</td>
<td>3.95</td>
<td>3.17</td>
</tr>
<tr>
<td>15</td>
<td>3.44</td>
<td>3.84</td>
<td>3.00</td>
</tr>
</tbody>
</table>

### Exhibit 7

**Variance analysis**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Chi-square</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>1356.10</td>
<td>0.00</td>
</tr>
<tr>
<td>Repeat (gender=1)</td>
<td>3</td>
<td>5.85</td>
<td>0.12</td>
</tr>
<tr>
<td>Gender</td>
<td>1</td>
<td>7.55</td>
<td>0.01</td>
</tr>
<tr>
<td>Residual</td>
<td>3</td>
<td>4.77</td>
<td>0.19</td>
</tr>
</tbody>
</table>

### ADDITIONAL READING


