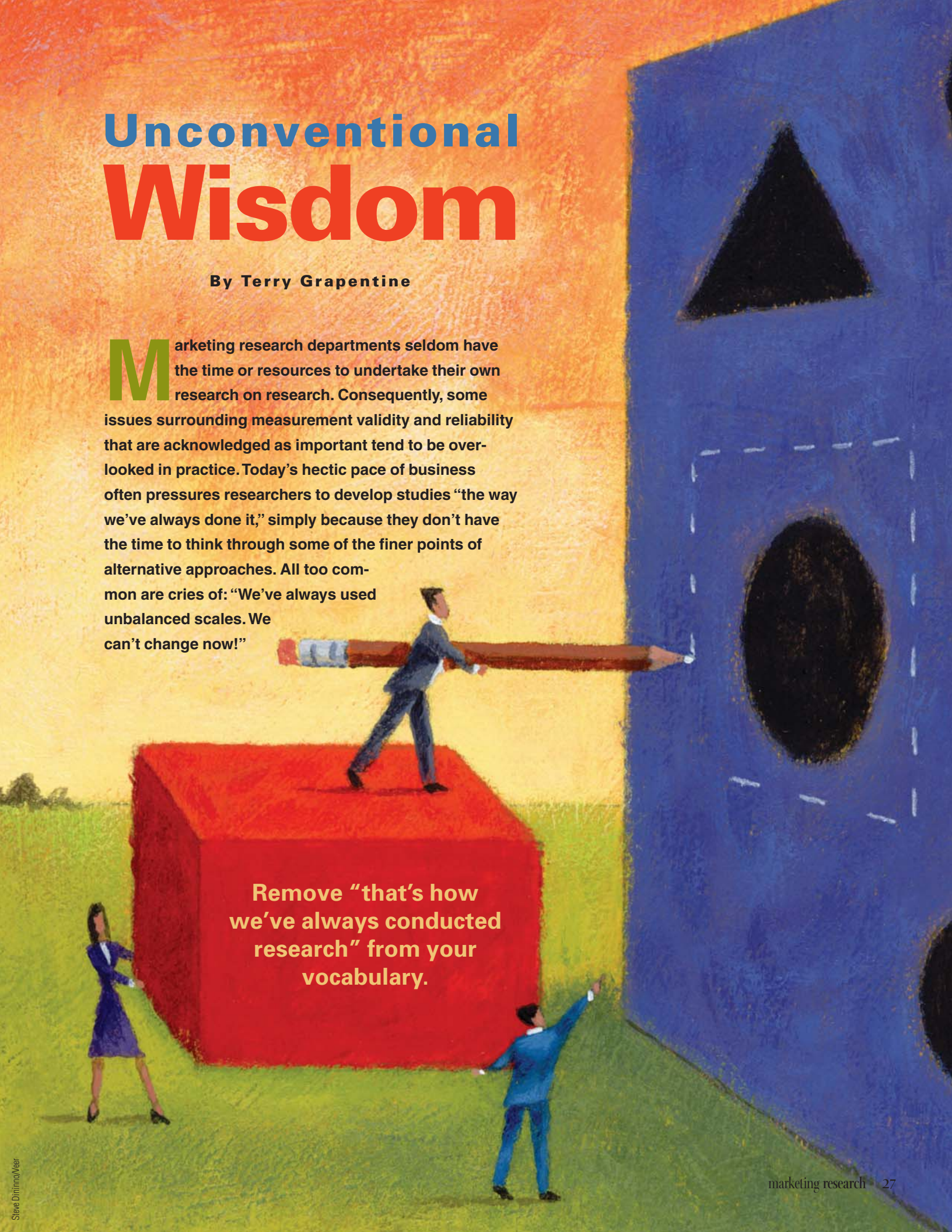
Unconventional Wisdom

By Terry Grapentine

Marketing research departments seldom have the time or resources to undertake their own research on research. Consequently, some issues surrounding measurement validity and reliability that are acknowledged as important tend to be overlooked in practice. Today's hectic pace of business often pressures researchers to develop studies "the way we've always done it," simply because they don't have the time to think through some of the finer points of alternative approaches. All too common are cries of: "We've always used unbalanced scales. We can't change now!"



Steve Diininno/Veer

Executive Summary

The pace of today's business environment often compels marketing research departments to conduct research the way they have done it historically, without pausing to scrutinize important issues that affect the quality of their products. This article examines several topics related to scale development and analysis, which can influence the validity and reliability of survey results. And it proposes that research departments periodically look at some of their "time-tested" methods to see whether they still pass muster.

This article offers a starting point to challenge the conventional wisdom in your research organization, by reexamining some commonly used scales and analytical techniques: scale anchor definitions, providing "don't know" responses, the statistical efficiency of summated scales, balanced vs. unbalanced scales, and factor analysis.

Clearly it's not a comprehensive list, but it's a start. When I first encountered these topics nearly 30 years ago, I felt consensus on them would be achieved by the 21st century. This has not happened, and I'm not sure why. Is it because the issues are complex, because their resolution is a function of the research context, or because marketing researchers are simply a contentious lot? Your letters to the editor might provide some insight.

Scale Anchors

Models that purport to predict a dependent variable measure, such as customer loyalty or purchase intention, additionally incorporate a set of predictor or independent variables. These independent variables measure a product's perceived performance on a set of attributes that are causally linked to the dependent variable. Various kinds of models—such as regression analysis, discriminant analysis, path analysis, or structural equation modeling—can be developed to examine the relationships between these independent and dependent measures.

How should the scale anchors for the independent variables in these models be constructed? Specifically, should the lowest point be labeled Extremely Dissatisfied and the highest point be labeled Extremely Satisfied?

Several years ago, my company had an opportunity to examine this question. We were presented with a data set in which a subset of product performance attributes—the independent variables—was measured twice. Both scales used a 1-10. The only difference was the scale anchors. One scale used 1 = Poor to 10 = Excellent. The other used 1 = Extremely Dissatisfied to 10 = Extremely Satisfied. The attribute ratings appeared on different pages of a self-administered, multipage questionnaire, and were far enough apart that we believe respondent answers to both sets of questions could be considered relatively independent of each another. The dependent variable of overall satisfaction with the product used a 1- to 10-point Extremely Dissatisfied/Extremely Satisfied scale.

Our hypothesis going into the analysis was that the variance in attribute ratings using the Poor/Excellent scale would be significantly greater than the variance in the attribute ratings of the alternative scale. That was because the Dissatisfied/Satisfied anchors bias respondent answers when they are used to measure independent variables. For example, if respondents are generally satisfied with a product, then they will tend to use the upper portion of this scale simply because they are satisfied with the product—not because of how a product performs on a specific attribute. This is a type of halo effect.

We tested our hypothesis by analyzing the data. Exhibit 1 presents analysis of the mean and variance statistics for nine attributes that were measured on the two different scales.

The data support the hypothesis. The average attribute variance for the Dissatisfied/Satisfied scale is 3.6, compared with 5.2 for the Poor/Excellent scale. As a consequence, the average attribute mean for the Dissatisfied/Satisfied scale (8.4) is significantly higher than that for the Poor/Excellent scale (7.6).

Measuring independent variables on a Dissatisfied/Satisfied scale is inappropriate when the dependent measure is itself a measure of satisfaction, or contains a satisfaction component. Theoretically, perceived product performance predicts satisfaction. Stating that independent variables measured on a satisfaction scale predict a global satisfaction measure is a bit of a tautology, like stating that retail sales predict gross national product.

Exhibit 1 Effects of scale anchors

	Scale A: 1 (poor) to 10 (excellent)		Scale B: 1 (extremely dissatisfied) to 10 (extremely satisfied)	
Attribute	**Mean**	**Variance**	**Mean**	**Variance**
A	8.0	4.6	8.3	4.0
B	6.8	6.7	8.3	4.0
C	6.9	5.9	8.4	3.3
D	7.2	5.9	8.4	3.7
E	8.1	4.4	8.4	3.6
F	8.3	4.2	8.7	3.0
G	7.6	5.5	8.2	4.0
H	7.7	5.0	8.5	3.5
I	7.8	4.5	8.6	3.6
Average	7.6	5.2	8.4	3.6

"Don't Knows"

Many situations arise in marketing research, in which respondents rate brands on a series of product attributes or answer a series of agree/disagree statements.

"Don't know" is sometimes a valid answer to such questions, but should we give respondents this option or force them to choose a positive or negative answer? There is no arguing that the two approaches can produce widely different answers! Consider Exhibit 2, regarding the question about charter schools asked in a recent Harvard University study (*National Public Radio/Kaiser/Kennedy School Education Survey*, National Public Radio Web site, www.npr.org).

If not given a "don't know" option, we've found that respondents do one of two things: Either they don't answer the question, or—when using a rating scale (say a 0- to 10-point scale)—they use the scale's midpoint (i.e., a 5 rating). Many also report a significant level of frustration and even irritability at not knowing what to do.

In most cases, consider providing a "don't know" response category. Using the scale midpoint for "don't know" can introduce significant measurement error into the data, affecting both simple statistics and parameter estimates of predictive models—as well as hiding important information from management.

Simple statistics such as the mean and median will be biased. For example, say respondents are using a 0- to 10-point scale to rate a product's performance, and the mean statistic on a given attribute is 6.2. Assume a significant number of respondents truly don't know how to rate a brand on that attribute and give the brand a 5 rating because "don't know" was not provided. This biases the true mean score downward. The computer shows the mean is 6.2. But in reality, among those who have an image of the brand on that attribute, the mean is 7.9.

Parameter estimates of predictive models will be biased as well. For instance, in the previous example, a significant number of such 5 ratings will artificially constrict a variable's variance. This results in a downward bias of that variable's importance in a regression model (or any other kind of advanced modeling technique, such as structural equation modeling). Regression results might indicate a given variable is not an important predictor of brand loyalty. But among those who have an image of a brand on that attribute, it is one of the most significant predictors of brand loyalty.

Finally, not including a "don't know" response option hides important information, when many respondents who truly don't have an image of a product on a given attribute use the scale midpoint as their answers. And that's especially important if your organization is investing significant resources to educate respondents about your product on that issue.

The only reason not to give respondents a "don't know" option: when exploratory research suggests that everyone should hold an opinion on the questions to be investigated. A thorough test of the survey instrument will provide insight on this issue.

So the next time you begin a research study and draft a questionnaire, be sure you've got "don't knows."

Summated Scales

In markets where it's becoming increasingly difficult to differentiate your product from the competition, increasing the statistical precision of your measures can be a competitive advantage. Summated scales offer this opportunity.

A summated scale is an index measure of a fundamental perceptual or attitudinal dimension of a consumer. Factor analysis provides insight into how to group attributes that reflect these underlying dimensions. For example, consider the two examples in Exhibit 3 on page 30. Assume that respondents used a 0- to 10-point scale—where higher numbers denote stronger agreement with the attribute—and the data are for one hypothetical respondent.

In these examples, the summated scale for each respondent serves as an index measure for the underlying dimension.

One interesting characteristic of the summated scale is that its variance is less than the variance of the individual items that constitute it. This increases the precision of your statistical tests. For example, Exhibit 4 on page 30 gives the standard deviation statistic for a summated scale and its items from an actual study.

The largest gap between the standard deviation of the summated scale and the standard deviations of the individual items that comprise it is for Attribute 2: a difference of .91. This smaller standard deviation for the summated scale increases the statistical precision of this measure compared with using only single items. For example, for the summated scale to be within .3 of the population mean (at the 95% confidence level), you would need a sample size of about 230. For Attribute 2, this sample size

Exhibit 2 Charter school study

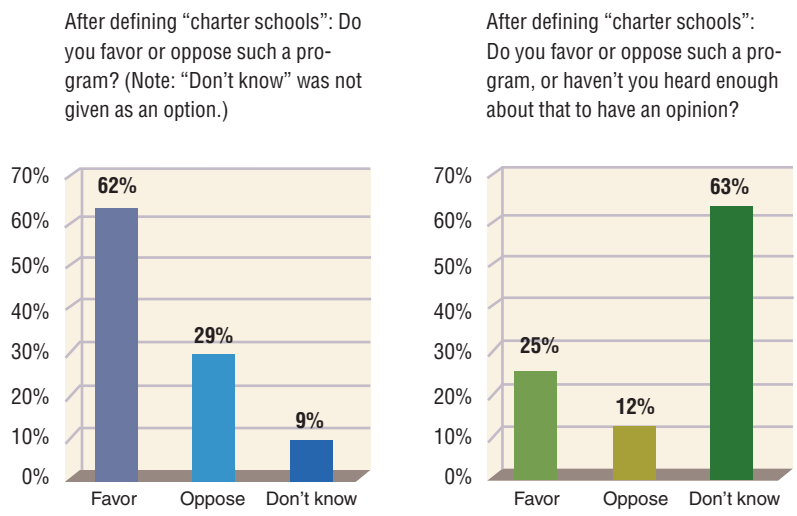


figure would be around 440. If you are paying upwards of $30 per interview, this can translate into significant dollars.

There are two additional advantages of summated scales. First, they are a more valid reflection of the underlying dimension being measured than any of the individual items that comprise it. It's like a math test in school: Any single question is a less valid measure of one's math skills than the entire battery of questions on the test. Second, using summated scales—instead of individual items—in regression modeling can reduce the effects of multicollinearity in the estimation of the model's parameters. Two pernicious effects of multicollinearity: Regression coefficients might be far from their true and unknown values, or their values might take on the wrong sign—the coefficients might take on negative values when theory or common sense suggests they should be positive.

Using summated scales increases the precision of your statistical tests. If there is a difference in brand image between competitive products, you are more likely to detect this difference with a summated scale than if you solely rely on the individual attributes of your survey.

Balanced Vs. Unbalanced Scales

Consider (1) a 0- to 10-point scale, a balanced scale whose midpoint is 5 (Example A), and (2) a 1- to 10-point scale, an unbalanced scale that does not have a midpoint (Example B).

Jum C. Nunnally and Ira Bernstein, who wrote the bible on psychological measurement (*Psychometric Theory*, McGraw-Hill, 1994), tell us that such scales reflect "how much of an attribute is present in an object." Scale values tell us how much of an attribute a consumer perceives a product (object) to possess, or—if attributes describe consumers—how much of an attribute consumers feel that they possess.

Exhibit 3 Summated scales

Perceptual dimension: perceived quality of a product		**Attitudinal dimension:** attitude toward assuming risk in an investment	
Attributes	**Ratings**	**Attributes**	**Ratings**
Lasts a long time	8	I will invest in the stock market in the future	2
Doesn't break easily	5	I'm willing to assume some risk to get above-average returns	4
Made of durable components	7	Most of my investments are in stocks	1
Won't scratch easily	7	Some of my investments are in emerging industries	3
Sum of ratings	27	Sum of ratings	10
Mean or "summated scale" = (8 + 5 + 7 + 7) / 4	6.8	Mean or "summated scale" = (2 + 4 + 1 + 3) / 4	2.5

Summated scales are like a math test in school: Any single question is a less valid measure of one's math skills than the entire battery of questions.

The amount of an attribute lies on a continuum from "very little/nothing" to "a lot," where "very little/nothing" is measured by the smallest number on a scale (0 or 1 in the previous examples) and "a lot" is measured by the largest number on a scale (10 in both examples). The operative word is continuum, for there is no theoretical reason to believe that this continuum is discontinuous—at least for many, if not all, of the attitude and belief attributes we measure in marketing research.

As an experiment, go outside and look at the sky. Use a scale from 0 to 10, where 0 denotes completely cloudy and 10 denotes crystal clear. On a crystal clear day, you'd give a rating of 10; on a completely cloudy day, you'd give a rating of 0. If you perceived that exactly 50% of the sky is clear and clouds obscure the other 50%, given the scale in Example A, you'd probably use the midpoint rating of 5.

But what if the scale did not have a midpoint? Using Example B, you might want to state 5.5, but be forced to give a rating of either 5 or 6—neither of which accurately reflects the amount of cloudiness that you perceive in the sky.

Unbalanced scales, therefore, introduce a slight amount of measurement error into the data. More measurement error means higher variances in survey measures. The result: You are less likely to detect a statistically significant difference

Exhibit 4 Standard deviations

Variable	Standard deviation
Attribute 1	2.71
Attribute 2	3.21
Attribute 3	2.71
Attribute 4	2.99
Summated scale	2.30

between two mean scores when in reality, such a difference exists—not a good thing, especially in tracking studies.

Fibbing Factor Analysis

Factor analysis helps researchers identify latent constructs by means of examining the relationships within a set of variables. These variables are only indicators of the constructs, because they cannot be directly observed.

In the financial services industry, "proactive service" is an example of a latent construct. It reflects the extent a financial institution takes initiative to serve the customer—as opposed to waiting for the customer to prod the institution to provide a new product or offer better service. One way of measuring this construct is by having respondents rate their primary financial institutions on a set of agree/disagree statements, such as the following:

- continually looks for better ways to serve its customers
- is one of the first to offer new products and services
- tailors its products and services to meet the needs of its customers
- keeps customers informed about new products and services

If these are good measures of proactive service, then they should load on the same factor.

True Love

In a study we did many years ago, respondents used a semantic differential scale to rate (1) what Valentine's Day means to them and (2) various placards that were to appear in a retail store during the week of Valentine's Day. Example attributes used in this exercise:

- emotional/not emotional
- tender/not tender
- fun/not fun

When we factor-analyzed data set 1, we denoted the first factor as romantic friendship, comprising the following attributes:

- emotional/not emotional
- romantic/not romantic
- affectionate/not affectionate
- passionate/not passionate
- tender/not tender
- caring/not caring

However, when we ran a discriminant analysis with the placards as the classification variables and the survey attributes as the predictor variables (data set 2), the analysis produced two functions: one that we labeled romantic friendship and the other that we labeled filial affection.

Variables loading high on the first factor of romantic friendship:

- emotional/not emotional
- romantic/not romantic
- passionate/not passionate

Variables loading high on the second factor of filial affection:

- affectionate/not affectionate
- tender/not tender
- caring/not caring

When using data set 1, which measured what Valentine's Day means to people, factor analysis combined these variables into one factor. With data set 2, in which respondents evaluated the advertising stimuli, discriminant analysis separated the variables into two discriminant functions.

Of course, factor and discriminant analyses perform different tasks. The former is looking for groups of variables that are intercorrelated. The latter is looking for variables that help us understand why two or more objects are perceived to be different.

In the Valentine's Day example, factor analysis fibbed, in a sense. In truth, the six semantic differential scales reflect two different, fundamental dimensions of Valentine's Day. All you had to do was look at the pictures being tested. One depicted a young couple strolling through the woods; the other depicted a grandfather with his granddaughter on his knee.

By separating the six variables into two groups of measures, the analyst can more validly examine how well advertising communicates to the consumer. More generally, with an appropriate data set, use both factor and discriminant analyses. Compare and contrast the results of the two methods for additional insight into the dimensions underlying your data.

Concluding Remarks

Marketing research departments are often agents of change for their internal clients. In fact, one could state that internal clients come to research departments to help promote change in the organization, by making new and better products. Yet, while doing new-product research for others, how often do research departments question their own methods/techniques or conduct their own research on research? This article touches on a few approaches affecting measurement validity and reliability, as a foundation for such a self-assessment. Does your department eschew balanced scales because it wants to force respondents to "take a side," or not give respondents a "don't know" response because they should have an opinion? At a minimum, consider the arguments. Do some research on research, to test some of the claims made here. Just don't let inertia or tradition be your master. ●

Terry Grapentine is president of Grapentine Co., a marketing research firm based in Ankeny, Iowa. He may be reached at terry@grapentine.com.