Measurement of Consumers' Perceptions of Product Quality, Brand Name, and Packaging: Candy Bar Comparisons by Magnitude Estimation

The magnitude estimation procedure provides a means to present consumer judgments about quality comparisons between brands on a linear, ratio-determined proportional scale that can be directly compared and validated. The procedure allows for a direct assessment of the added value with which the brand name and package endow product quality. A comparison of 14 candy bars is presented for illustration.

The findings from a convenience sample of consumers are presented to illustrate the magnitude estimation approach for the assessment of perceived product quality. This procedure can be applied to a small sample of consumers and is seen as a useful quantitative application to otherwise qualitative focus group studies. Thirty respondents evaluated blind-rated quality, the brand name, and packaging of 14 chocolate candy bars in a pilot example and application of the technique.

The magnitude estimation approach allows for a valid interpretation of consumers' judgments about perceived product quality. The results reported here provide a demonstration of a quantitative value-added assessment of the components of product quality. They apply only to respondents from the illustrative sample, however, and are presented solely to demonstrate the usefulness of the procedure.

The isolation and measurement of the components of product quality are important to marketing and product management, to researchers interested in understanding consumer behavior, and to market researchers for the meth-
odology necessary to establish an applicable quality research program.

The Problems of Quantifying Quality

Many of the problems that arise in the researcher’s assessment of product quality are related to the tools used to measure its conceptualization. Jacoby and Olson (1984) note that the critical shortcoming in the assessment of consumer perceptions of product quality is the lack of an appropriate standardized measurement. Furthermore, most previous research findings cannot be generalized easily because of a lack of consistency across studies and the varied methods and research designs used to identify attributes comprising product quality (Gardner 1971; Monroe and Dodds 1988).

The Concept of Perceived Product Quality

The concept of perceived product quality is defined to include the consumer’s response to the entire evoked set of judgments about quality comparisons among competitive brands. This concept differs from the current practice of defining product quality on the basis of a multidimensional list of product attributes to make the construct operational. Typically, consumers are asked to evaluate the importance of a given set of attributes for a product category. They then evaluate the degree to which each brand has those attributes. Their estimate of quality is obtained by multiplying the attribute’s importance by the rated evaluation of the attribute, then summing across all of the attributes to obtain an overall product quality score (Fishbein and Ajzen 1980).

This procedure distorts the construct because the concept of quality depends on a predetermined list of product attributes used to define it. The list of attributes selected to represent the construct may be incomplete, inaccurate, and distorted, regardless of the rigors in the attribute selection process. Jaccard, Brinberg, and Ackerman (1986) found little consistency among six methods common in the assessment of attributes and their importance.

Advocates of an aggregate one-dimensional concept of quality note the potential for error in the use of attributes to define product quality (Curry and Faulds 1986). However, any one-dimensional aggregate measure of quality raises the issue of scale validity. How does the researcher know the consumers’ impressions are accurate if only one measure is used? The validity of a construct is an issue of measurement rather than definition. A valid measurement tool must be selected to make operational a comprehensive construct for the meaning of product quality.

What Is a Magnitude Scale and How Does It Work?

The magnitude scale is considered a ratio scale of psychological and perceptual stimuli. Respondents assign numbers to objects (or brands, etc.) so that ratios between the assigned numbers reflect ratios among the objects on the criterion being tested (Dillon, Madden, and Firtle 1987). Some recent marketing applications of this scale include aggregate measures of emotions from advertisements (Neibeker 1984), new product testing (Moskowitz 1985), and quality measure-
ments (Lavenka 1989). Crask and Fox (1987) and Teas (1987) support the validity of the magnitude scaling methods when aggregate data are analyzed.

Stevens (1946) marshalled the standard convention for the assignment of numerals to scales, commonly known as the nominal, ordinal, interval, and ratio measurement scales. He also endorsed a fifth measurement scale called the "magnitude scale," which is sensitive to logarithmic data and satisfies the criterion of additivity and proportionality.

Stevens' (1975) Power Law governs the respondents' sensory impressions of stimulus magnitude, as all human sensory responses have known logarithmic constants. The respondents' task of drawing lines (converted to millimeters) and their assignment of numbers to stimuli both have the same human sensory exponent of 1.0. These two sensory measures offer common data collection procedures and simplified instructions for the subjects to follow (Crask and Fox 1987). (See instructions in the Appendix.)

Figure 1 is a plot of logarithms of the geometric means of the line length and numeric estimates obtained from the re-

**Figure 1**

**CALIBRATION SLOPE OBTAINED BY PLOTTING LINE LENGTH VERSUS NUMERICAL ESTIMATES OF THE TWO SENSORY MEASURES (log-log scale)**

\[ y = -2.43 + 0.99x \]
respondents' comparisons of overall quality. The log-log scale of the two measurement responses is expected to produce approximately a straight line with a slope of 1.0. Validity is established to the extent that relationship holds on the basis of the Power Law. This model yields a slope of .993 (within the probability of the theoretical slope) and a correlation coefficient of .94 (substantiates the linear relationship).

The appropriate measure of central tendency for magnitude data is the geometric mean. Geometric means have the advantage of being expressed as percentage differences. Measurement comparisons translated into proportional differences facilitate the manager's direct interpretation of the relevancy and clarity of the data. Current measurement provides only ordinal data derived from the semantic differential (i.e., 1-to-7 scale) and interpreted by way of statistical inference.

Research Design Strategy for Magnitude Assessment

The study reported here incorporates a repeated-measures design in which each respondent is measured multiple times to provide ratings of competitive brands of chocolate candy bars. The repeated-measures design is widely employed in social science and can be used to remove otherwise hidden and unmeasurable systematic sources of variability (Cook and Campbell 1979). It also allows for an assessment of the causal direction and logical order of product quality evaluation, based first on the intrinsic blind-rated quality, then other extrinsic components such as packaging or brand name. The consumers' responses to brand equity (Farquhar 1989), as well as package equity, are isolated and can be assessed for the added value they bring to the product (Leuthesser 1988).

The chocolate candy bar was selected as a convenient illustration of the components of product quality. A wide selection of competitive brands is available for comparison, as well as varied intrinsic physical differences of quality. In addition, all candy bars share the same reference price of about 45 cents, which minimizes any possible price-quality judgments that may have a predetermined influence on consumers' perceptions (Monroe 1973). Two cases each of 14 conveniently selected candy bars were obtained from a Chicago-based distributor. The distributor provided coding dates ensuring freshness within two weeks for all 14 brands.

A convenience sample of 30 respondents age 18 to 25 was chosen for a pilot demonstration of the technique. A sample of heavy users (10 or more bars a week) was selected to profile the typical consumer of confectionary products. However, the results are only for illustration and are not representative of all consumers.

The guidelines from an updated review of Stevens' (1975) research in magnitude scaling, along with Lodge's (1981) validation approach, were followed to obtain the magnitude estimation measurements. The study design and application are described in the technical appendix.

Analysis of Test Results

The high correlations for each of the three tests ($R^2 = .95, .91$, and $.94$) indicate that the subjects were making consis-
tent responses in their judgments. Simple regression analysis was used to obtain the correlation of the geometric means of line length (converted to millimeters) and numeric estimation for each of the 14 brands. Each subject’s response was entered in log form (base 10) and the responses were summed to obtain the arithmetic mean of the logs for each stimulus. Then each mean was exponentiated (raised to the power of 10) to obtain the geometric mean.

The confidence intervals were based on the correlation from each of the two sensory measures. The empirical exponent was accepted when within the 95% confidence limits constructed around the theoretical exponent. The Student’s $t$ of 2.048 with 12 degrees of freedom was used to formulate the confidence limits around the obtained empirical exponents. Table 1 shows that the empirical betas are all within 95% of the theoretical beta, indicating a valid psychophysical application of the magnitude estimation procedure based on Steven’s Power Law.

### Interpretation of the Magnitude Estimation Results

Discussion of results is confined to general interpretation of the data derived from the magnitude scale. The specific findings are obviously limited to a small convenience sample of respondents from Chicago. Preferences and tastes can differ with the selection of products for comparison, types of consumers, regional differences, and the selection and pre-

### Table 1

**TEST RESULTS FOR BLIND-RATED QUALITY, BRAND NAMES, AND PACKAGE COMPONENTS FOR PERCEIVED PRODUCT QUALITY**  
*(N = 30 respondents)*

<table>
<thead>
<tr>
<th>Test 1</th>
<th>Test 2</th>
<th>Test 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Blind-Rated Product Quality</strong></td>
<td><strong>Product With Brand Names</strong></td>
<td><strong>Product, Brand Names, and Package</strong></td>
</tr>
<tr>
<td><strong>Name</strong></td>
<td><strong>Geometric Mean</strong></td>
<td><strong>% Rating</strong></td>
</tr>
<tr>
<td>Snickers</td>
<td>72.2</td>
<td>100.0</td>
</tr>
<tr>
<td>Heath</td>
<td>61.3</td>
<td>84.8</td>
</tr>
<tr>
<td>Skor</td>
<td>59.1</td>
<td>81.9</td>
</tr>
<tr>
<td>Milky Way</td>
<td>56.1</td>
<td>77.7</td>
</tr>
<tr>
<td>Oh Henry</td>
<td>51.2</td>
<td>70.9</td>
</tr>
<tr>
<td>(Average)</td>
<td>50</td>
<td>—</td>
</tr>
<tr>
<td>Butterfinger</td>
<td>47.2</td>
<td>65.3</td>
</tr>
<tr>
<td>3 Musketeers</td>
<td>45.7</td>
<td>63.3</td>
</tr>
<tr>
<td>Chocolate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pay Day</td>
<td>45.6</td>
<td>63.2</td>
</tr>
<tr>
<td>5th Avenue</td>
<td>45.4</td>
<td>62.9</td>
</tr>
<tr>
<td>Almond Joy</td>
<td>43.8</td>
<td>60.6</td>
</tr>
<tr>
<td>Butternut</td>
<td>41.1</td>
<td>56.9</td>
</tr>
<tr>
<td>Bounty</td>
<td>36.8</td>
<td>51.0</td>
</tr>
<tr>
<td>Clark</td>
<td>36.4</td>
<td>50.0</td>
</tr>
<tr>
<td>Good Stuff</td>
<td>33.0</td>
<td>45.7</td>
</tr>
</tbody>
</table>

$R^2 = .95^a$  
$\text{Beta} = 1.003^b$

$R^2 = .94^a$  
$\text{Beta} = 1.023^b$

\[ R^2 = .95^a \]  
\[ \text{Beta} = 1.003^b \]

\[ R^2 = .94^a \]  
\[ \text{Beta} = 1.023^b \]

$^aR^2$ is the product-moment correlation between the geometric means obtained from the two measurement scales of numeric estimation and line length.

$^b$All of the empirical betas are within the .95 confidence limits of the theoretical beta of 1.00.
sentation of quality components. The results from the three tests are reported in Table 1 for illustration of the magnitude scale.

Test 1 Results

The highest rated brand is Snickers. Its geometric mean of 72.2 is used as the percentage base with which all other judgments are compared. The brand with the lowest perceived quality is Good Stuff, which is perceived to have 2.2 times less quality than the leading brand. The Snickers product should be judged as the model or standard of blind-rated sensory taste quality. A geometric mean of 50 represents the "average" in perceived (blind) quality. Five bars are perceived to be above average and nine bars below average in blind-rated quality.

Heath is perceived to have 15% less quality than Snickers (61.3/72.2). Heath and Skor bars rate surprisingly high on quality. Both are toffee bars with real milk chocolate. Respondents perceived a 3.5% difference in quality (59.1/61.3) between the two bars. Good Stuff, a regional (Minneapolis) brand unknown to the respondents, is lowest in blind-rated quality (33).

Test 2 Results

Snickers remains the highest rated brand and Heath is second. However, the 15% perceived difference found in test 1 becomes a 26% difference when brand name is made known. (Note the increase in the geometric means from 72.2 to 83.4). The difference can be related to the equity of the Snickers brand name. This finding suggests that Heath should build the quality of its brand name to remain competitive with Snickers.

Butterfingers, Three Musketeers, and Milky Way all show increases in their previous above-average ratings with their brand names. However, these brands lose ground because of the large increase in the Snickers brand name rating.

Test 3 Results

When the Snickers package is considered in the evaluation of product quality, its rating increases from 83.4 to 85. Snickers also improves its distance from Heath to a 28.6% difference (60.7/85). Heath now rates fourth in product quality, which suggests that Heath packaging needs updating.

Second and third place ranks are replaced with Three Musketeers and Milky Way when the package is considered. The three top leaders are Mars, Inc. products. Mars’ packaging affords added value to consumers’ perceptions of overall product quality. Only the Mars bars had consumer open-code dating indicating the freshness of the product. This source of consumer information may have led to the increase in package equity, which can be isolated and evaluated in future studies.

Useful Applications of the Magnitude Estimation Approach for Product Quality and Brand Equity Assessment

Employing two different sensory tests provides powerful criterion validity that goes well beyond simple reliability
Concurrent validity is found for line length and numeric estimation.

Measures obtained from standard category scales. The concurrent validity of the two measurements, line length and numeric estimation, for product quality assessment is indicated by the high product moment correlations. The face criterion validity of the concept of product quality is addressed as the respondents' judgments follow known stimulus ratios in support of Stevens' Power Law. The measurement results are based on the assumption of full information, both objective and subjective, and isolated for each quality component by means of the multiple-measures design.

The multiple-measures design of the study was used to illustrate how the components of product quality are isolated on the basis of blind-rated quality. The magnitude estimation approach provides sufficient explanation of aggregate product quality, whereas the multiple-measures design allows the isolation of conditions necessary for a complete and logically consistent evaluation of the components of product quality. Assessment of the effects of the extrinsic components of product quality can be based on the consistency of a blind-rated base measure of intrinsic quality. The equity that the brand name and package bring to the product can be derived from an index based on the magnitude estimation results. Other components of product quality, such as store name and design quality, also could be isolated and assessed for their contribution to aggregate perceptions of product quality.

The valuation of intangible marketing assets (e.g., brand name, packaging, etc.) is made difficult because of the lack of consistent and reliable measures of consumer preferences. Current financial assessments of brand equity, for example, are often made with such market information as market share, market trends, brand stability, and other trade-generated data. These measures involve the observation of overt after-the-fact behavior, which limits managers' ability to understand and predict marketing behavior. The isolation and measurement of consumer perceptions of marketing-related variables by means of direct interviews gives the manager primary data with which to direct mass marketing programs and strategies. The leverage that a brand name and package provide over other comparable products can be assessed by way of a consumer perspective.

The study attempts to address the priority need to quantify perceived product quality. It suggests a research program to assess the components of product quality necessary to evaluate the alternatives for marketing management decisions. These components represent real and exploitable assets of the firm and every effort to quantify them is valuable. The scope of this study can be extended to evaluate the relationship of product quality to buyer types, products, price assessment, service satisfaction, and other moderating variables.

Appendix

The complete study involved a 35-minute personal interview. A five-minute calibration exercise was used to acquaint the respondents with assigning numbers to lines and drawing lines to assigned numbers. Three tests involving different levels of product information lasted a total of 10 minutes each. Specific instructions and the sequence of tests follow.
Practice Exercises

You will be answering questions using unique rating procedures. I ask you to complete a practice exercise so that you can become familiar with the rating scales. First, here is a sheet of paper with a series of lines drawn on it. Kindly note that some of the lines are longer than the first line (A is 50 millimeters long) and some are shorter. The first line is your reference. Let us give it the number 50. Your task is to say how much longer or shorter the lines are compared to the first line by giving each line a number compared to 50. The longer a line appears to be compared to the reference line, the bigger the number you will give it compared to 50. The shorter a line compared to the first line, the smaller the number you should give it compared to 50. Here is a sheet with numbers on it. Note the numbers in the above left of this page. Your task is to draw a line (free hand) according to how much longer or shorter the lines are compared to the number given. The first line is our reference number. We have given it the number 50. All you need do is to draw a line, longer or shorter, than the line given you at 50.

Instructions: Test 1

Products were unwrapped and placed on paper plates coded A to N. The candy bars were randomized to avoid order effects. After five minutes of handling and tasting all candy bars, the respondent was asked to assign a number (with 50 as average) to each bar according to the question:

"HOW DO YOU RATE THE QUALITY OF EACH CANDY BAR GIVEN ONLY THE INFORMATION IN FRONT OF YOU?"

After finishing this assignment, the respondent was asked to draw a line for each candy bar with the same question. A 50-millimeter line was provided as an average reference line. Respondents could not refer back to any previous assignment.

Instructions: Test 2

After the respondents had tasted the products, the brand names (written in block letters on 3 × 5" cards) were attached to each plate. Respondents then were asked to assess the quality and brand name with each of the two sensory tests.

"HOW DO YOU RATE THE QUALITY AND BRAND NAME OF EACH CANDY BAR GIVEN ONLY THE INFORMATION IN FRONT OF YOU?"

Instructions: Test 3

Respondents were presented a packaged candy bar and asked to assess the following question with each of the two sensory tests:

"HOW DO YOU RATE THE QUALITY, BRAND NAME, AND PACKAGE OF EACH CANDY BAR GIVEN ONLY THE INFORMATION IN FRONT OF YOU?"

References


