

Multi-item Scale Usage in Marketing Journals: 1980 to 1989

Gordon C. Bruner II

Southern Illinois University

Paul J. Hensel

University of New Orleans

The use of multi-item scales reported in six marketing journals during the 1980s is analyzed. The analysis replicates some aspects of the Churchill and Peter (1984) study and extends the examination to issues not reviewed previously in marketing. The database for the study is unique in that it attempts to incorporate every instance of scale usage from the defined domain. Among the findings is that the use of multi-item scales increased substantially during the 1980s but the reliability of those scales was not different from earlier periods. A majority of scales have their origins in marketing and nearly half of all scales were used to measure consumer behavior constructs.

As a discipline advances toward being a science it should periodically look back on its past to determine the progress it has made. There have been several studies in the last few years that have reviewed the state of research in published marketing studies. For example, there were the articles in the *Journal of the Academy of Marketing Science's* special issue on the State of the Art in Marketing Research (1988) and several more in the recent 20th Anniversary issue of the journal (1992).

However, a comprehensive examination of one of our most commonly used measurement tools, the multi-item scale, has not been made in nearly a decade.¹ The intent of this study is to provide such a review by replicating some aspects of the Churchill and Peter (1984) study as well as

examining some other issues that have not been reviewed previously.

BACKGROUND AND PURPOSE

Two serious marketing measurement problems in the 1960s and 1970s were the over usage of single-item measures and the lack of scales developed specifically for use in marketing (Kassarjian 1971, p. 415; Peter 1981, p. 138). However, the 1980s presented a different problem; while it might be assumed that progress was made in terms of scale development and usage, there is little objective information about how far the field progressed in that usage.

There have been only a few previous efforts to understand and describe various aspects of scale usage in marketing (e.g., Churchill and Peter 1984; Peter 1979, 1981; Peter and Churchill 1986; Edris and Meidan 1990). The study reported here is most similar to that by Churchill and Peter (1984). Although it was not the purpose of this study to copy all aspects of that seminal work, there are enough similarities that some comparisons can be made. Specifically, those few scale or study characteristics that were found to have significant associations with scale reliability (scale length, number of response intervals, sample size) as measured by Cronbach's (1951) coefficient alpha were re-examined in this study.

Additional aspects not previously examined in a marketing context were also incorporated into this review. The journal and year of publication were noted in order to evaluate scale usage over time and how the scholarly publications can be distinguished based upon the scales reported in them. Further, the scales in this study have been separated into three groups that have been identified as the most studied in marketing: consumer issues, advertising issues, and organizational issues (Malhotra 1988).² This will provide some

idea of the quantity and quality of scales across construct areas.

The first six hypotheses presented below are directional in nature since the literature provides a rationale for expecting certain findings. Following that, there are three null hypotheses that will be tested as well.

- H1.** The number of scale items and level of internal consistency are positively correlated.
- H2.** The number of response alternatives and level of internal consistency are positively correlated.
- H3.** Sample size and internal consistency are negatively correlated.
- H4.** The number of scales published in scholarly marketing journals has increased over time.
- H5.** Average scale reliability has increased over time.
- H6.** The majority of scales used in published marketing research were developed outside of marketing.
- H7.** There is no difference between the major marketing journals in the mean internal consistency of the scales published in them.
- H8.** There is no difference in the proportion of scales that have been used to measure consumer, advertising, and organizational constructs.
- H9.** There is no difference between the three major construct groups in their mean levels of scale internal consistency.

METHODOLOGY

The multi-year task of compiling and describing the scales used in published marketing research was formidable. Limits had to be imposed to make the task achievable in some reasonable time period. The two main limits placed on the domain of review were the number of journals and years of publication. Six journals were ultimately selected: *Journal of the Academy of Marketing Science*, *Journal of Advertising*, *Journal of Advertising Research*, *Journal of Consumer Research*, *Journal of Marketing*, and *Journal of Marketing Research*. During the 1980s they were among the most well known and respected journals in the field (Browne and Becker 1985; Fry, Walters, and Scheuermann 1985; Luke and Doke 1987).

As noted above, previous studies have mainly examined aspects of scale usage in the 1960s and 1970s. Therefore, it seemed reasonable to limit the domain in the present study to the most recent decade, that is, 1980 to 1989.

Further, to determine the total number of scales reported and the change in scale usage over time, an attempt was made to include every scale in the database that met certain

TABLE 1
Deciles for Metric-level Scale Characteristics^a

Decile	Characteristic			
	Number of Items	Number of Scale Points	Internal Consistency	Sample Size
1	2	5	.59	83
2	3	5	.66	114
3	3	5	.71	145
4	4	5	.75	180
5	4	6	.79	216
6	5	7	.82	299
7	6	7	.85	379
8	7	7	.88	456
9	10	7	.91	639
10	40	11	.98	1792

^aThe first decile indicates the value of each characteristic at the 10 percent level read from their individual frequency distribution tables. The fifth decile can be viewed as the median for each of the scale characteristics. At the other extreme, the tenth decile denotes the highest values for each of the characteristics observed in the data.

criteria. First, scales had to have two or more items. Second, a minimum amount of information had to be known about each scale, particularly with regard to psychometric properties and item content. This information generally was gathered from within the articles themselves, other published sources, or personal correspondence.

The coding task was tedious due to the sheer number of scale applications and the difficulty of gathering the data, particularly when the information was not clearly stated in the articles. Inter-coder reliabilities were estimated to be above .90 for all variables except scale origination (.72).³ Adjustment in coding instructions improved reliability substantially but caution is still advised for any conclusions that might be drawn regarding scale origin.

In total, over 2,400 articles, comments, and research notes were examined for measures that met the stated criteria for inclusion in the database. Seven hundred fifty uses of codable scales were found in 181 different articles. Therefore, even though this study is not a census of all scale usage in marketing research, it does represent an attempt to include as many scales from the defined domain as met the specified criteria. Given this, the data would appear to be a large and comprehensive sample from which a reasonably good view of scale usage and quality from the decade of the 1980s can be judged. Frequency distributions in the form of deciles for several of the key metric-level variables are presented in Table 1.

FINDINGS

H1. The internal consistency of a scale as estimated by coefficient alpha is determined by the number of items composing a scale and the degree of covariation among them (DeVellis 1991, p. 86). If the average correlation between items is low then alpha will be more affected by the number of items than when the level of intercorrelation is high. Past research has provided empirical evidence that the number of

TABLE 2
Correlations, Means, and Standard Deviations for Metric-level Scale Characteristics

<i>Characteristic</i>	<i>Year Published</i>	<i>Scale Length</i>	<i>Scale Points</i>	<i>Reliability</i>	<i>Sample Size</i>
Scale length	-0.13 ^a				
Scale points	-0.02	-0.20 ^a			
Reliability	0.12 ^a	0.21 ^a	0.14 ^a		
Sample size	0.04	-0.08 ^b	-0.13 ^a	-0.21 ^a	
Mean	1986 ^c	5.96	6.06	0.77	327.58

^a $p < .001$.

^b.001 < $p < .05$.

^cMedian year of publication.

scale items has a positive relationship with internal consistency (e.g., Churchill and Peter 1984). As indicated in Table 2, the evidence was consistent with previous research and supported **H1**.

H2. Reliability typically increases as the number of scaling intervals increase (Churchill and Peter 1984; Cox 1980). This is due to the tendency for items with a larger number of points to possess larger covariances and variances than the same items with fewer response alternatives (Martin 1973, 1978). In support of this hypothesis, a significant though low correlation was found between the number of scale points and internal consistency. When the number of items was controlled for, a partial correlation of .19 resulted ($p = .0001$).

H3. One might expect that sample size has some influence on internal consistency but it might not be evident as to what direction that influence might take. Evidence of a negative relationship has been found previously (Churchill and Peter 1984) and **H3** was confirmed in the present study. Specifically, a correlation of $-.21$ was found.

H4. This hypothesis was endorsed as indicated by the data provided in Table 3. It shows that there was a six-fold increase of scale usage during the 1980s, from 24 applications in 1980 to 152 in 1989. Although the increase was not monotonic, the general trend was upward.

H5. Even though the reliability of scales used in the 1980s ranged from a very low .28 to .98, the mean was a

respectable .77. This range and mean are strikingly similar to what was reported by Churchill and Peter (1984) for an earlier period. So, while there was a slight association ($r = .12$) between year of publication and internal consistency, it appears to be inappropriate to conclude that reliability increased steadily throughout the decade or that it was substantially greater than it was for the previous two decades.

H6. Past evidence has indicated that a large percentage if not the majority of scales used in marketing have been borrowed from other fields (e.g., Kassarian 1971; Peter 1981). Despite this, **H6** was rejected since the data here indicated that only about 18 percent of the scales were developed outside of marketing (Table 3). In contrast, about 27 percent were original to the study in which they were reported, and 26 percent were borrowed from previous marketing studies. The origin of the remaining scales reviewed could not be determined due to a lack of information.

H7. There is little doubt that our scholarly journals are perceived to vary in their quality and status (Browne and Becker 1985; Fry, Walters, and Scheuermann 1985; Luke and Doke 1987). What is lacking is some objective confirmation of these perceptions. One objective criterion that can be used to compare journals is mean reliability of the measures used in them. In this study mean internal consistency was found to range from .72 for the *Journal of the Academy of Marketing Science* to .79 for both the *Journal of Consumer Research* and the *Journal of Marketing Research*. **H7** was

TABLE 3
Origination Status, Scales Published, and Reliability By Year

<i>Year of Publication</i>	<i>Original (%)</i>	<i>Borrowed/ Marketing (%)</i>	<i>Borrowed/ Non-Marketing (%)</i>	<i>Unknown (%)</i>	<i>Number Published</i>	<i>Mean Reliability</i>
1980	45.8	4.2	41.7	8.3	24	.80
1981	41.9	16.1	29.0	12.9	31	.68
1982	18.2	52.3	0	29.5	44	.73
1983	32.1	39.6	17.0	11.3	53	.75
1984	25.6	20.5	25.6	28.2	39	.78
1985	19.2	30.4	14.4	36.0	125	.74
1986	33.0	13.6	38.6	14.8	88	.78
1987	28.0	18.7	6.7	46.7	75	.81
1988	24.4	36.1	8.4	31.1	119	.76
1989	25.7	19.7	17.8	36.8	152	.78
Mean ^a	26.8	26.0	17.6	29.6	75	.77

^aThe percentages in the four origination status cells on this row are based on the entire data set.

rejected given that an ANOVA indicated the differences in journal means were statistically significant ($F = 7.02, p = .0001$). Perhaps this reflects a difference in the constructs being studied in the various journals and, as indicated by **H9** below, measurement of some constructs appears to be more reliable than that of others.

H8. This hypothesis was rejected since nearly half (47%) of all scale uses were found to involve consumer constructs. Forty-one percent of the scales related to organizational issues and the remainder of the scales (12%) were used to assess advertising constructs. If one subsumes advertising scales under consumer-related constructs then clearly the broad area of consumer behavior has received the majority of scale users' attention.

H9. An ANOVA indicated that there was significant variation in the mean reliabilities of scales belonging to the three major construct groups leading to the rejection of **H9** ($F = 8.23, p = .0003$). Specifically, the advertising group was found to have scales with the highest mean internal consistency (.82), followed by the organizational and consumer groups (both approximately .76). This suggests that the constructs of interest to marketing researchers may vary in their difficulty of measurement.⁴

DISCUSSION

The findings presented here should be interpreted in light of several limitations. First, although the scales came from six top marketing-related journals, many other journals as well as proceedings were not reviewed. It is possible, therefore, that scales published in sources that for whatever reason are not as popular or have lower status would have characteristics and relationships different from those examined. Second, as extensive as the database was, it lacked scales for which sufficient information could not be obtained. There is no reason to believe, however, that as a group such scales were any different from what was included. Third, nothing should be inferred from this study regarding single item scales. While they were used extensively during the 1980s, caution should be exercised in assuming anything about their nature from what is reported here regarding multi-item scales. Finally, this study did not directly address issues of scale validity. Attempts to examine the different types of validity were frustrated by the inconsistent usage of terminology and the varying interpretation of results that have been reported in the literature.

Among the clear findings of this study are that the three significant relationships reported by Churchill and Peter (1984) were confirmed here as well: internal consistency was positively related to the number of scale items and points but was negatively associated with sample size. The combined impact of these associations on internal consistency must be kept in perspective, however. A regression analysis indicated that the number of items, the number of scaling points, and sample size produced an R -squared of only .11 ($p = .0001$) when used to predict alpha. Therefore, although these factors should receive serious attention when developing and using individual measures, they appear to explain only a limited amount of the variance in scale reliability for the population as a whole. Other factors such as

the nature of the construct, the nature of the respondents, and the average intercorrelation of scale items may ultimately explain more variance.

Brevity of measures is an important issue to marketing researchers due to the time and space limitations typically encountered in data collection (Malhotra 1992, p. 384). Indeed, it was found here that 53 percent of the scales had between two and four items. Since reliability can suffer with short scales and adding more items can be a problem, consideration should be given to increasing the number of response alternatives per scale item. If deemed appropriate for other reasons, more response alternatives could yield a greater range of scores that in turn may make a scale more sensitive and reliable (DeVellis 1991, pp. 64, 65; Nunnally 1978, pp. 595, 596). A rule of thumb is that the number of potential scores for a summated scale (highest possible score minus lowest possible score) should be at least 10 if not greater than 20 to minimize the effect of attenuation (Martin 1973, 1978). At the other extreme, there appear to be diminishing returns in measure sensitivity when too many scale items or response alternatives are used (Green and Rao 1970).⁵

While the general level of scale internal consistency in the domain was acceptable, we were struck by the number of times researchers would attempt to justify low reliabilities ($\alpha < .7$) by citing the older edition of Nunnally (1967, p. 226) rather than the more recent edition (1978, pp. 245, 246) in which higher levels are recommended. Low reliabilities should be a concern to researchers because they decrease the statistical power of tests. In other words, along with other factors, the probability of detecting significant relationships of particular magnitudes improves when scales of higher reliability are used (DeVellis 1991, p. 32).

A balanced approach to internal consistency is advocated here, however. Although the errors of low reliability have typically been the greatest concern, there is growing concern over the blind pursuit of high reliability as well (e.g., Kline 1986). If scale developers merely select those items from a pool that maximize reliability, then the result is what Cattell referred to as a "bloated specific" (1978, p. 289). Therefore, reasonably high reliability is a worthy goal but should not be pursued at the expense of adequately capturing the essence of some construct's domain (validity).

It is clear to us after personally examining each of the scales that new measures have been constructed too many times when acceptable ones have already been reported in the literature. In contrast to what Peter (1981) observed a decade ago, there are literally hundreds of scales available to assess a wide variety of marketing-related constructs. Therefore, the challenge facing most scale users today is not so much having to construct good scales from scratch or having to borrow and adapt measures from other fields; it is a matter of doing proper reviews of the marketing literature before questionnaires are prepared. The task has now been made easier with the publication of two compilations of marketing scales (Bearden, Netemeyer, and Mobley 1993; Bruner and Hensel 1992).

The findings reported here concerning scale reliability are not meant to address the question of overall journal quality. But, it is suggested that average scale reliability is one criterion that could be considered for use in tandem with

other criteria to more objectively compare journal quality than has occurred so far. Other criteria that could be considered are rejection rate, age of journal, subscription base, and the number of citations in other journals as well as in citation indices.

Although this study was able to determine that consumer-related constructs have been scaled more than advertising or organizational constructs in the domain examined, a more specific question could not be answered: how many *distinct* constructs have had scales developed for their measurement? This study confirms what Perreault (1992, p. 371) pointed out in a recent review: marketing scientists have put considerable effort into developing scales for the measurement of a wide variety of constructs. We cannot be specific at this point, however, about what that variety is nor can we list the constructs for which good measures exist and those which still require scale developers' attention. One could get an idea of the variety of constructs by examining the scale handbooks referred to above. A comprehensive inventory of scaled constructs may have to wait until a hierarchical schemata of marketing constructs, possibly resembling what has been done in psychology (Walker 1991), is first produced.

NOTES

1. The term scale is used throughout this paper to refer to multi-item psychometric measures for which scores are computed for respondents by adding or averaging numeric responses from individual scale items. The term is not used here for single item measures nor at the other extreme for inventories with multiple "subscales." Further, the characteristics of scale usage are based on the number of individual scale uses, not the number of different and distinct scales themselves.
2. The term "organizational" is used broadly here to refer to constructs relating to sales management, personal selling, channels, and other constructs involving marketing management that are not directly related to consumers or advertising.
3. Inter-coder reliabilities for nominal data were calculated using the formula offered by Perreault and Leigh (1989). Reliabilities for ratio level data were calculated by dividing the number of agreements among the judges by the number of observations they made.
4. Although the entire database has not been broken down into fine construct categories, preliminary evidence indicates clearly that the level of reliable measurement varies widely from one construct to another.
5. While admitting that their recommendations were preliminary, Green and Rao (1970) were quite specific in concluding from their study that little is gained by increasing the number of scale items beyond eight or the number of points beyond six.

REFERENCES

- Bearden, William O., Richard G. Netemeyer, and Mary F. Mobley. 1993. *Handbook of Marketing Scales*. Newbury Park, CA: Sage.
- Browne, William G. and Boris W. Becker. 1985. "Perceptions of Marketing Journals: Awareness and Quality Evaluations." In *Proceedings of the American Marketing Association Educators' Conference*. Chicago: American Marketing Association, 149-154.
- Bruner, Gordon C., II and Paul J. Hensel. 1992. *Marketing Scales Handbook: A Compilation of Multi-Item Measures*. Chicago: American Marketing Association.
- Cattell, Raymond B. 1978. *The Scientific Use of Factor Analysis in Behavior and Life Sciences*. New York: Plenum.

- Churchill, Gilbert A., Jr. and J. Paul Peter. 1984. "Research Design Effects on the Reliability of Rating Scales: A Meta-Analysis." *Journal of Marketing Research* 21 (November): 360-375.
- Cox Eli P., III. 1980. "The Optimal Number of Response Alternatives for a Scale: A Review." *Journal of Marketing Research* 17 (November): 407-422.
- Cronbach, Lee J. 1951. "Coefficient Alpha and the Internal Structure of Tests." *Psychometrika* 16 (September): 297-334.
- DeVellis, Robert F. 1991. *Scale Development: Theory and Applications*. Newbury Park, CA: Sage.
- Edris, Thabet A. and A. Meidan. 1990. "On the Reliability of Psychographic Research: Encouraging Signs for Measurement Accuracy and Methodology in Consumer Research." *European Journal of Marketing* 24 (3): 23-41.
- Fry, Elaine Hobbs, C. Glenn Walters, and Lawrence E. Scheuermann. 1985. "Perceived Quality of Fifty Selected Journals: Academicians and Practitioners." *Journal of the Academy of Marketing Science* 13 (Spring): 352-361.
- Green, Paul E. and Vithala R. Rao. 1970. "Rating Scales and Information Recovery: How Many Scales and Response Categories to Use?" *Journal of Marketing* 34 (July): 33-39.
- Kassarjian, Harold H. 1971. "Personality and Consumer Behavior: A Review." *Journal of Marketing Research* 8 (November): 409-418.
- Kline, Paul. 1986. *A Handbook of Test Construction: Introduction to Psychometric Design*. New York: Methuen.
- Luke, Robert H. and E. Reed Doke. 1987. "Marketing Journal Hierarchies: Faculty Perceptions, 1986-1987." *Journal of the Academy of Marketing Science* 15 (Spring): 74-78.
- Malhotra, Naresh K. 1988. "Some Observations on the State of the Art in Marketing Research." *Journal of the Academy of Marketing Science* 16 (Spring): 4-24.
- . 1992. "Shifting Perspective on the Shifting Paradigm in Marketing Research: A New Paradigm in Marketing Research." *Journal of the Academy of Marketing Science* 20 (Fall): 379-387.
- Martin, Warren S. 1973. "The Effects of Scaling on the Correlation Coefficient: A Test of Validity." *Journal of Marketing Research* 10 (August): 316-318.
- . 1978. "Effects of Scaling on the Correlation Coefficient: Additional Considerations." *Journal of Marketing Research* 15 (May): 304-308.
- Nunnally, Jum C. 1967. *Psychometric Theory*. New York: McGraw-Hill.
- . 1978. *Psychometric Theory*. Second Edition. New York: McGraw-Hill.
- Perreault, William D., Jr. 1992. "The Shifting Paradigm in Marketing Research." *Journal of the Academy of Marketing Science* 20 (Fall): 367-375.
- Perreault, William D., Jr. and Lawrence E. Leigh. 1989. "Reliability of Nominal Data Based on Qualitative Judgments." *Journal of Marketing Research* 26 (May): 135-148.
- Peter, J. Paul. 1979. "Reliability: A Review of Psychometric Basics and Recent Marketing Practices." *Journal of Marketing Research* 16 (February): 6-17.
- . 1981. "Construct Validity: A Review of Basic Issues and Marketing Practices." *Journal of Marketing Research* 18 (May): 133-145.
- Peter, J. Paul and Gilbert A. Churchill, Jr. 1986. "Relationships Among Research Design Choices and Psychometric Properties of Rating Scales: A Meta-Analysis." *Journal of Marketing Research* 23 (February): 1-10.
- Walker, Alvin, Jr. 1991. *Thesaurus of Psychological Index Terms*. Arlington, VA: American Psychological Association.

ABOUT THE AUTHORS

Gordon C. Bruner II is Associate Professor of marketing at Southern Illinois University. He is Director of the Office of Scale Research and has written the *Marketing Scales Handbook* along with Dr. Hensel. His research has been published in the *Journal of the Academy of Marketing Science*, *Journal of Marketing*, *Journal of Retailing*, *Journal of Marketing Education*, and many other journals and conference proceedings.

Paul J. Hensel is Professor of marketing at the University of New Orleans. His research has emphasized measurement and advertising issues as well as ethical and social issues. He has published in journals such as the *Journal of the*

Academy of Marketing Science, *Journal of Advertising*, *Journal of Health Care Marketing*, and *Journal of Business Ethics*. He also co-authored the *Marketing Scales Handbook* with Dr. Bruner.