

Re-Inquiries

How to Construct a Test of Scientific Knowledge in Consumer Behavior

JOHN R. ROSSITER*

Scientific knowledge in consumer behavior is defined as consisting of consumer behavior structural frameworks or models (microtheories) and well-supported empirical generalizations in various areas of consumer behavior (microfindings). This re-inquiry first examines a pioneering attempt to develop a test of scientific knowledge in consumer behavior, the Armstrong Test. The problems with that test are instructive in revealing threats to validity in test construction and analysis. Second, detailed steps are proposed for constructing a comprehensive, valid test of scientific knowledge in consumer behavior. Such a test should be useful for assessing the consumer behavior knowledge held by business educators, consultants, managers, market researchers, and business students.

How should a test of scientific knowledge in consumer behavior be constructed? Armstrong, in this journal (September 1991), pioneered the difficult task of constructing a test of this type. But the test produced disturbing results that demand a re-inquiry. According to this test, herein referred to as the Armstrong Test, experts in consumer behavior—consumer behavior academics and marketing managers—demonstrated no more scientific knowledge of consumer behavior than did tenth and eleventh grade high school students in honors English classes. That is, the experts scored no better on the test than intelligent laypersons, and thus the test failed to demonstrate known-groups validity. Similarly disturbing were the mean knowledge scores obtained by the three groups, which were not much better than chance. According to this test, none of the groups knew

much about consumer behavior, which is a very questionable outcome.

The present endeavor begins by identifying the constituents of scientific knowledge in consumer behavior. The use of “in” is deliberate because the reference, like the reference of the Armstrong Test, is to scientific knowledge in the consumer behavior domain (knowledge about the how and why of consumer behavior) as contrasted with descriptive knowledge “of” consumer behavior (knowing what typical consumers think and do). See Hoch’s (1988) test for the latter purpose. Following this conceptual discussion, a critical analysis of the Armstrong Test is offered that explains why that test produced such disconcerting findings. Using the conceptual discussion, and learning from the earlier attempt, the re-inquiry concludes with detailed recommendations for the construction of a valid test of scientific knowledge in consumer behavior.

WHAT IS SCIENTIFIC KNOWLEDGE IN CONSUMER BEHAVIOR?

To appreciate what a valid test would cover—and to understand what the Armstrong Test actually did cover—it is instructive to consider what “scientific knowledge in consumer behavior” would comprise. In the broader domain of marketing, Rossiter (2001, 2002b) has proposed that declarative knowledge (knowing what, as distinct from pro-

*John R. Rossiter is research professor of marketing in the School of Management, Marketing and Employment Relations, University of Wollongong, Wollongong NSW 2522, Australia (john_rossiter@uow.edu.au) and professor of marketing on the visiting faculty in the Department of Marketing Management, Rotterdam School of Management, Erasmus University, The Netherlands. The author thanks Scott Armstrong for providing his test materials; Elizabeth Cowley, Grahame Dowling, Sandra Jones, and especially Terry Shimp and Ale Smidts for their advice on a lengthier previous version of the article; and is indebted to *JCR* editors David Glen Mick and Kent Monroe as well as three *JCR* reviewers for their constructive suggestions for preparation of the present version.

cedural knowledge, or knowing how, which in any case presumes accurate declarative knowledge) takes five forms: (a) concepts, which are definitions of key constructs; (b) structural frameworks, which relate concepts as descriptive models and are thus descriptive microtheories; (c) empirical generalizations, which are well-established beliefs about “if, then” correlations between concepts, in other words, generalized findings; (d) strategic marketing management principles, which are causal microtheories prescribing dynamic “if, do” marketing actions; and (e) market research principles, which are causal microtheories prescribing “if, use” problem-specific research techniques.

Applied to knowledge in consumer behavior as a subfield of marketing, strategic marketing management principles, form *d* above, would drop out, as these principles are in the seller behavior subfield. Also excluded would be market research principles, form *e* above, because these are a specialized subfield of marketing knowledge. The relevant forms of knowledge in consumer behavior would therefore be

1. Consumer behavior concepts—for example, involvement, repeat-buying.
2. Consumer behavior structural frameworks—for example, Howard’s (1977) three-stage models of extensive problem solving, limited problem solving, and routinized response behavior; the FCB advertising planning grid (Vaughan 1980).
3. Consumer behavior empirical generalizations—for example, Huff’s (1964) retail gravitation law; Ehrenberg’s (1988) duplication-of-purchase law.

Among these forms, when one thinks of “scientific knowledge” in a field, one is usually thinking of empirical generalizations—established findings in that field (see Hubbard and Lindsay 2002; *Marketing Science* 1995; Uncles 2002). However, knowledge of structural frameworks is similarly important. Structural framework models such as Howard’s three-stage models, Fishbein-type multiattribute attitude models (Ahtola 1975), the Elaboration Likelihood Model (Petty and Cacioppo 1986), various grid models, and checklist models such as the extra “3P’s” that consumers look for in services are useful knowledge contributions to consumer behavior. It could be reasoned, therefore, that the most important forms of knowledge to test in consumer behavior are structural frameworks (microtheories) and empirical generalizations (microfindings). Concepts, the most basic form of knowledge, are subsumed in structural frameworks and empirical generalizations. However, it is essential that the concepts in structural frameworks and empirical generalizations are clearly translated into everyday language (otherwise the test of knowing the structural frameworks or empirical generalizations is confounded by lack of prior knowledge of the technical terms in which they are stated).

WHAT DID THE ARMSTRONG TEST MEASURE?

On the surface, the Armstrong Test might appear to be a test of knowledge of empirical generalizations in consumer behavior. The test presented hypotheses from 20 empirical studies published in the *Journal of Consumer Research*, together with a brief description of the purpose and method of each study, and asked respondents to predict the outcomes of the tests of those hypotheses by indicating whether each hypothesis was true or false. Thus, apparently, if a respondent knew the empirical generalization relevant to that study, then the respondent could use this knowledge to predict the specific outcome. For example, one of the studies used in the Armstrong Test was Swinyard and Coney’s (1978) study of the effects of advertising and personal telephone contact on voting behavior; if the respondent knew that advertising tends to have a small effect on behavioral intentions for high-involvement choices and a large effect for low-involvement choices, then the respondent would be able to correctly predict those two outcomes for the study.

However, the stated purpose of the Armstrong Test was much more ambitious than this. Armstrong set out to test people’s knowledge of theories of consumer behavior, from which generalizable findings might be inferred. Armstrong’s unedited conclusion, given in his original working paper (1989, pp. 18, 19), was that “the field [of consumer behavior] has not been progressing as a science . . . [in that] the theories available to academics do not help them in predicting the phenomena in their major field of interest.” This is too idealistic; the respondent would have to know only the relevant empirical generalization, not the theory behind it, presuming there is one, to answer the questions.

But did the Armstrong Test achieve even the more limited aim of testing people’s knowledge of empirical generalizations in consumer behavior? The method by which the test was constructed, resulting in the procedure that participants were asked to follow, suggests that this aim was not achieved. For the test, participants were provided with a two-page introduction to each study which outlined, in nontechnical language, the purpose of the study, sample, method, and measures. The introduction deliberately excluded mention of prior studies and their findings, which existed for 19 of the 20 studies (Armstrong 1991, p. 254). Following the introduction was a list of the hypotheses of the study, likewise stated in nontechnical language. For the response task, participants had to indicate whether each of the study’s hypotheses was true or false, that is, whether the hypothesis was confirmed or disconfirmed. They could alternatively answer “do not understand.”

The Question (Item) Content

The Armstrong Test can be criticized firstly and most fundamentally for its question (item) content. Provision of the introductions, without mention of prior studies, is likely to have converted the participants’ task from the intended purpose of judging whether the hypothesis was true in gen-

eral (an empirical generalization) to one of judging whether the hypothesis was true in particular, that is, whether the finding was obtained in this one particular study given the sample and method used in that study. In other words, it is likely that respondents were led to make conditional judgments about the truth or falsity of specific findings. The introductions, which were completely specific, made it unlikely that respondents would draw on knowledge of a generalization to predict a particular finding.

If participants made conditional judgments about the results of just the study described to them, this would eliminate the advantage that consumer behavior academics might have in their knowledge of previous studies on the topic and also eliminate the advantage that consumer behavior academics and marketing managers might have over high school students in their knowledge of relevant empirical generalizations. The difficulty of making these conditional judgments would also account for the low scores on the test, which were little better than chance, or guesswork.

The Criterion for Correct Answers

A further count against the claims that the Armstrong Test was measuring people's knowledge of empirical generalizations in consumer behavior, let alone of the possible theories behind them, is the fact that the correct answer used to score each item on the test was not the truth of the empirical generalization (e.g., the truth of whether, in general, advertising does have a large effect on behavioral intentions for low-involvement choices) but rather whether the generalization held or not in that particular study. In other words, the truth of the specific finding was the criterion for correct answers, not the truth of the generalization.

As Kluger and Tikochinsky (2001) have pointed out, a test of a hypothesis (a specific operationalization) cannot refute an abstract generalization (an empirical generalization) and neither, for that matter, can the failure of an empirical generalization, which is one derivation from a theory, refute that theory. That is, the hierarchy in science is theory → empirical generalization → hypothesis (a particular operationalization of an empirical generalization). A theory is generative, that is, it spawns a number of empirical generalizations (Eysenck 1997). Each empirical generalization can be tested (measured) in numerous ways, not all of which are valid (see Rossiter 2002a). Correct answers on the Armstrong Test were thus at the lowest scientific level and did not test knowledge of consumer behavior theories, as intended, and did not test knowledge of empirical generalizations.

Technical Criticisms of the Analysis

A technical criticism of the analysis in the Armstrong Test is that the instructions led respondents to believe that the distribution of true and false answers was 50 : 50 (Armstrong 1989, p. 14), as is usual on true-false tests, that is, that the base rate of true being correct was 50%. In actuality, 74% of the hypotheses were true (Armstrong 1991, p. 253), so that respondents using a strategy in accordance with the

stated base rate of correct answers would, on average, get 24% wrong due to overguessing of false. The realistic maximum score obtainable on the test would then be 76%, not 100%, a ceiling score closer to the mean scores that were observed.

A further technical criticism, which makes the observed scores even more debatable, is that directional outcomes reported in the original studies were counted in the test as supporting the hypotheses. The stricter criterion of statistically significant outcomes should have been used (one-tailed at the .05 level since the hypotheses made directional predictions). Directional outcomes allow the dubious practice of treating, say, a mean difference of +1% in favor of the hypothesis as supporting the hypothesis, and of 0% or -1% as rejecting it, or of a minuscule +.01 correlation between two variables in the hypothesized direction as proving the hypothesis true. Lack of statistical significance of a particular finding makes it less probable that a generalized finding would be true (Hagen 1997; Nickerson 2001).

Explaining the Poor Results

It has already been suggested that consumer behavior academics, and also marketing managers, would have no advantage on the Armstrong Test if it tested particulars rather than empirical generalizations. Moreover, with respect to the poor performance by consumer behavior academics, the results from those academics who recalled having read the particular study in the original were excluded, thus making it less likely that those remaining could call on the relevant empirical generalization.

Attesting to the difficulty of the task that the Armstrong Test required of respondents were the high refusal rates. Of 100 consumer behavior academics approached, only 20 responded, and two of these were excluded because they said they had read most of the studies, and a further two were excused because they said they did not understand all the hypotheses, leaving a final academic sample of $n = 16$. Of the 100 marketing practitioners from the AMA's mailing list who were invited to participate, only 13 responded, and one of these was excused because he or she said they did not understand the task itself, leaving a final practitioner sample of $n = 12$. The high school students, on the other hand, were virtually forced to participate, in class, giving a student sample of $n = 43$. In the statistical analysis of the results, an attempt was made to compensate for these small n 's by aggregating across hypotheses as well as respondents (each respondent received about 26 of the 105 hypotheses from the 20 studies). This is not a valid procedure because the correct answers to the hypotheses within a study (there were approximately five hypotheses from each of the five studies given to each respondent) would not be independent.

In summary, the Armstrong Test cannot be regarded as a valid test of scientific knowledge in consumer behavior. The Armstrong Test gauged people's ability to predict specific findings given the particular method of the study; it did not test people's knowledge of consumer behavior empirical generalizations or of theories in consumer behavior.

There are plausible reasons why the test produced such strange results and why consumer behavior academics performed so badly on the test.

WHAT WOULD BE A VALID TEST?

The construction of a valid test of scientific knowledge in consumer behavior requires several carefully planned and executed steps. First, the forms of consumer behavior knowledge and their content domains must be determined. Second, test items need to be generated, with the concepts in these items translated into nontechnical language. Third, appropriate and unambiguous answer formats must be provided. Fourth, a small set of experts in each consumer behavior subdomain must ratify the items by providing a correct-answer profile. Fifth, the items and the answer alternatives must be pretested with a moderate-sized sample of laypersons representing the lowest consumer behavior education population to whom the test might be administered, to pretest for clarity of both the items and the answer formats. Sixth, the test can then be administered to groups of persons differing in formal consumer behavior education. Seventh, to score the test, answers should be compared with the best-answer profile provided by the experts. Steps 1–4 are elaborated below.

Item Forms

In terms of the forms of consumer behavior knowledge to be tested, it would be acceptable to limit these to structural frameworks (microtheories) and empirical generalizations (microfindings). As proposed previously, consumer behavior concepts, which are the elemental form of consumer behavior knowledge, would be given in the test items as lay translations rather than being tested separately. One reason for this is that a test of scientific jargon, in consumer behavior as in any other discipline, would artificially discriminate in favor of those people who have had formal education in the field. Another reason is that unless these concepts are translated, answers to the test items covering structural frameworks and empirical generalizations will be confounded by indeterminacy of whether failure on the item was due to lack of understanding of the concept descriptions or lack of the knowledge expressed therein.

Item Content

The content of the items would be selected by polling experts—ideal for this task would be the *JCR* editorial board and the journal's reviewers. Experts in each subdomain of consumer behavior would be polled. These subdomains could be the same as the content/subject areas used by this journal for potential reviewers to indicate their areas of expertise: "advertising; aesthetic/hedonic consumption; . . . word of mouth/opinion leadership."

The experts would be asked to provide four types of items—without regard, at this stage, for the degree of empirical support for the item. The items would consist of

- Useful structural frameworks or models in that subdomain
- Popular but less useful, even discredited, structural frameworks or models
- Empirical generalizations about consumer behavior in that subdomain (although stated as generalizations, they would be quite specific as to content, as in the example of advertising's influence on choice behavior in low-involvement and high-involvement decision situations)
- Mythical empirical generalizations (common misconceptions)

False-knowledge items, namely, questionable frameworks or models and common misconceptions about consumer behavior, are necessary types of items to be included on the test. These items are necessary theoretically because the eradication of false knowledge is as important as the inculcation of true knowledge. Also, such items are needed to ensure an even distribution of true and false answers on the test. Tests of popular misconceptions have been used extensively in psychology to assess people's scientific knowledge of the field, beginning with the test developed by McKeachie (1960) and formalized by E. Vaughan (1977) in her 80-item Test of Common Beliefs. Gardner and Hund (1983), using a sample of 20 of the most extremely falsely believed items, found that these items discriminated significantly, as expected, between psychology faculty and nonpsychology faculty.

Answer Format

The next step needing explanation is the selection of an answer format for the items. In the Armstrong Test, the answer alternatives were "true," "false," and "do not understand." The do not understand answer alternative should not be necessary with well-constructed items (not if technical terms in the items have been translated into everyday language and the item wording has been pretested for clarity). But the "mostly true–mostly false" or simple "true–false" answer format that is typically used for knowledge tests (e.g., Armstrong 1991; Gordon, Kleiman, and Hanie 1978; and Vaughan 1977) is not accurate. This can be illustrated by considering items of degree and items of degree and direction. For an item of degree, consider the empirical generalization that "personality shows only a small relationship with brand preferences." An answer of "mostly true" for this item could refer ambiguously to alternative interpretations, including reference to most, though not all, specific personality traits showing a small relation; most studies of personality traits and brand choice revealing a small relation; or even that the average relationship is not quite as large as "small." An answer of "mostly false" is also uninformative, as it could refer variously to the respondent's belief that the relationship is a lot smaller than "small," that is, zero; or that the relationship is indeed a lot larger, which of course are two very different answers reflecting two very different interpretations of the state of being false. Items that involve both degree and direction are

still more problematic. For instance, how could a respondent appropriately answer the item, "Recent new car buyers are less likely to subsequently pay attention to ads for the make and model of car they have just bought"? "Mostly false" or simply "false" would capture neither the fact that the relationship is typically in the opposite direction nor the average strength of that relationship. As can be shown with the Armstrong Test (and with the tests in psychology), the true-false answer format cannot accurately capture the respondent's state of knowledge.

Instead, the test constructor must work much harder and devise concrete answer alternatives specific to each item. For each item, one of the answer alternatives would be the best answer given by experts (see below) and the others would be plausible distracters. The provision of five answer alternatives would keep the proportion of correct answers obtainable by guessing at a fairly low level (about 20%). The alternatives should exclude the global "none of the above" or "all of the above." For instance, alternatives for the car-buyer item and advertising might be: "very much less likely," "less likely," "neither less nor more likely," "more likely," or "much more likely." For items that refer to structural frameworks or models, the answer alternatives would include one answer listing the correct components and four others listing erroneous but plausible components. For instance, for Huff's retail gravitation law, the correct (best) answer would state that patronage is directly proportional to retail selling space and inversely proportional to travel time. The distracter answers might refer to plausible erroneous variations of Huff's determinants of patronage, such as number of product categories and travel distance.

Answer Ratification (Scoring)

In the next step, experts in the respective subdomains of consumer behavior would be engaged to ratify the best answers and provide an answer profile for the items pertaining to their area of expertise. Knowledge in the social sciences consists of best beliefs held by experts (Rossiter 2001, 2002b) as there are no absolute truths. The experts doing the ratification should be subdomain experts (specialists). Experience from another ambitious study is instructive here. Miner (1984) assessed 32 established theoretical frameworks from the domain of personnel management in terms of their scientific validity. For these assessments, he intended to rely on the judgments of the 35 knowledgeable scholars who nominated the frameworks. However, he found that he could not use a full count of the panel for the assessments because his own extensive analysis of the literature revealed that many of the experts were not fully up-to-date—except with regard to the evidence pertaining to the several frameworks in their area of specialization. There is little doubt that the distribution of current scientific knowledge would be similarly uneven among academics in consumer behavior.

Scores on the test would thus be derived on the basis of an expertly ratified, best-answers profile. Scores would be zero-one for each item, regardless of the closeness of the four alternative distracters per item, as the purpose is to seek

precise best beliefs rather than to encourage even close misconceptions. With the many subdomains of consumer behavior knowledge, with several microtheories, microfindings and misconceptions and erroneously believed findings, a total test comprising 250 to 300 items would be required. This is not a test of ability but rather knowledge, so a smaller sample of items cannot be substituted. However, it is possible that shorter yet comprehensive subtests of subdomain knowledge would find applications.

CONCLUSION

Consumer behavior, as an established if evolving discipline, should have available a compendium of scientific knowledge in its domain. This collection would have to go beyond textbooks and be drawn from an assessment of many articles, by experts. By using the conceptual approach and method described in this article, a valid test of consumer behavior knowledge can be constructed. The test would be useful for assessing the consumer behavior knowledge acquired by business educators, consultants, managers, market researchers, and business students.

[David Glen Mick served as editor and Kent B. Monroe served as associate editor for this article.]

REFERENCES

- Ahtola, Olli T. (1975), "The Vector Model of Preferences: An Alternative to the Fishbein Model," *Journal of Marketing Research*, 12 (February), 52–59.
- Armstrong, J. Scott (1989), "Do Theories of Consumer Behavior Aid Predictions? Predictive Validity as a Measure of Scientific Progress," working paper, Wharton School, University of Pennsylvania, Philadelphia, March.
- (1991), "Prediction of Consumer Behavior by Experts and Novices," *Journal of Consumer Research*, 18 (September), 251–256.
- Ehrenberg, A. S. C. (1988), *Repeat-Buying*, 2d ed., London: Griffin.
- Eysenck, Hans J. (1997), "Personality and Experimental Psychology: The Unification of Psychology and the Possibility of a Paradigm," *Journal of Personality and Social Psychology*, 73 (December), 1224–1237.
- Gardner, Rick M. and Renee M. Hund (1983), "Misconceptions of Psychology among Academicians," *Teaching of Psychology*, 10 (February), 20–22.
- Gordon, Michael E., Lawrence S. Kleiman, and Charles A. Hanie (1978), "Industrial-Organizational Psychology: Open Thy Ears O House of Israel," *American Psychologist*, 33 (October), 893–905.
- Hagen, Richard L. (1997), "In Praise of the Null Hypothesis Statistical Test," *American Psychologist*, 52 (January), 15–24.
- Hoch, Stephen J. (1988), "Who Do We Know: Predicting the Interests and Opinions of the American Consumer," *Journal of Consumer Research*, 15 (December), 315–324.
- Howard, John A. (1977), *Consumer Behavior: Application of Theory*, New York: McGraw-Hill.
- Hubbard, Raymond and R. Murray Lindsay (2002), "How the Emphasis on 'Original' Empirical Marketing Research Impedes

- Knowledge Development," *Marketing Theory*, 2 (December), 381–402.
- Huff, David L. (1964), "Defining and Estimating a Trading Area," *Journal of Marketing*, 28 (July), 34–38.
- Kluger, Avraham N. and Jonathan Tirochinsky (2001), "The Error of Accepting the 'Theoretical' Null Hypothesis: The Rise, Fall, and Resurrection of Commonsense Hypotheses in Psychology," *Psychological Bulletin*, 127 (May), 408–423.
- Marketing Science* (1995), "Special Issue on Empirical Generalizations in Marketing," 14 (No. 3, Pt. 2).
- McKeachie, William J. (1960), "Changes in Scores on the Northwestern Misconceptions Test in Six Elementary Psychology Courses," *Journal of Educational Psychology*, 51 (December), 240–244.
- Miner, John B. (1984), "The Validity and Usefulness of Theories in an Emerging Organizational Science," *Academy of Management Review*, 9 (April), 296–306.
- Nickerson, Raymond S. (2001), "Null Hypothesis Significance Testing: A Review of an Old and Continuing Controversy," *Psychological Methods*, 5 (June), 241–301.
- Petty, Richard E. and John T. Cacioppo (1986), *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*, New York: Springer.
- Rossiter, John R. (2001), "What is Marketing Knowledge? Stage I: Forms of Marketing Knowledge," *Marketing Theory*, 1 (September), 9–26.
- (2002a), "The C-OAR-SE Procedure for Scale Development in Marketing," *International Journal of Research in Marketing*, 19 (December), 1–31.
- (2002b), "The Five Forms of Transmissible, Usable Marketing Knowledge," *Marketing Theory*, 2 (December), 369–380.
- Swinyard, William R. and Kenneth A. Coney (1978), "Promotional Effects on a High- versus Low-Involvement Electorate," *Journal of Consumer Research*, 5 (June), 41–48.
- Uncles, Mark (2002), "From Marketing Knowledge to Marketing Principles," *Marketing Theory*, 2 (December), 345–353.
- Vaughan, Eva D. (1977), "Misconceptions about Psychology among Introductory Psychology Students," *Teaching of Psychology*, 4 (October), 138–141.
- Vaughan, Richard (1980), "How Advertising Works—a Planning Model," *Journal of Advertising Research*, 20 (September/October), 27–33.