



A verbal rating scale to measure Japanese consumers' perceptions of product quality

A verbal rating scale

179

Boris Bartikowski
Euromed Management, Marseille, France

Katsuyuki Kamei
Faculty of Informatics, Kansai University, Osaka, Japan, and

Jean-Louis Chandon
Université Paul Cézanne Aix-Marseille (IAE-CERGAM), Puyricard, France

Received October 2008
Revised August 2009
Accepted August 2009

Abstract

Purpose – This paper aims to investigate whether verbal rating scales are viable formats for attitude measurement through an application to Japanese consumers' product quality perceptions.

Design/methodology/approach – The article notes theoretical differences between Likert-based and Thurstone approaches to attitude measurement. The paper illustrates a Thurstone scale development process.

Findings – The new scale possesses nomological validity; it correctly predicts how consumer ethnocentrism relates to product quality evaluations for brands in different competitive situations.

Practical implications – The convenient, ready-to-apply verbal rating scale can measure Japanese consumers' perceptions of product quality. The article also offers survey researchers some practical guidance for developing their own verbal rating scales.

Originality/value – Verbal rating scales are rarely found in existing literature. This study sheds light on a frequently overlooked measurement scale format for measuring attitudes.

Keywords Ethnocentrism, Japan, Consumer behaviour, Behaviourally-anchored rating scales, Individual perception

Paper type Research paper

1. Introduction

Measuring peoples' opinions or attitudes on questionnaires using measurement scales is a common practice in all fields of management. To facilitate survey respondents' evaluation task, researchers sometimes use rating scales with verbal descriptions of the scale anchors, referred to as Thurstone or verbal rating scales. For example, stakeholders might evaluate the trustworthiness of a firm's press announcements according to gradually differing expressions such as "sure and certain," "quite true," "almost reliable," "rather implausible," and "false." Verbal rating scales offer several advantages for attitude measurement, in that they are convenient to use and help reduce questionnaire length. In some cases, they may even improve measurement validity, according to the results of recent psychometric studies. However, the advantages of Thurstone scales frequently get overlooked, and the scale format rarely is employed in current literature. Therefore, this research attempts to shed light on Thurstone scales as a viable alternative to attitude measures. To this end, we discuss



The authors gratefully acknowledge the financial support of Kansai University, which awarded a research fellowship to realize this project. The authors especially acknowledge Teichi Kawata, former president of Kansai University, Shinichi Kitani, Dean of its Faculty of Informatics, and staff at its Division of International Affairs who supported the authors' research at Faculty of Informatics, Takatsuki Campus, Kanasi University in June and July in 2007.

Asia Pacific Journal of Marketing
and Logistics
Vol. 22 No. 2, 2010
pp. 179-195
© Emerald Group Publishing Limited
1355-5855
DOI 10.1108/13555851011026935

the advantages and theoretical differences of a Thurstone, as opposed to the dominant Likert, approach to attitude scaling. Furthermore, we illustrate our Thurstone scale development procedure, which results in a new verbal rating scale for measuring Japanese consumers' perceptions of product quality. The new scale achieves appropriate psychometric qualities for interval-level measurement; we also demonstrate the reliability and nomological validity of the scale.

Customer-perceived quality represents a key determinant of sustainable business success. A firm's product or service quality, relative to that of its competitors, is "the most important single factor affecting a business unit's performance" (Buzzell and Gale, 1987, p. 7). Higher quality perceptions lead, in the short run, to increased profits due to premium prices and, in the long run, to effective business growth, involving both market expansion and market share gains. More specifically, quality and reliability have long been considered integral to Japan's industry and production processes, as illustrated by the slogans for Sharp Electronics, "Quality First in Heart and Mind," and Fujitsu, "Quality built-in, with cost and performance as prime considerations." The focus of the Japanese Kaizen philosophy is to improve quality constantly and recognize the possibility of changing all functions of a business, from manufacturing to management, from the CEO to the assembly line workers (Masaaki, 1986). Kaoru Ishikawa, one of the originators of total quality control management in Japan, outlined as early as 1968 that "quality comes first, not short-term profits" (Pecht and Boulton, 1995, p. 116). Customer-driven quality remains one of the seven core criteria required by the Japan Quality Award (see www.jqac.com).

In this context, we review the Likert approach to attitude measurement, then discuss its advantages, as well as those of the Thurstone approach. To illustrate the Thurstone scale development procedure, we conduct a series of three studies that feature item generation, scale development, and scale validation. Finally, we conclude with some theoretical and managerial implications.

2. Likert and Thurstone approaches to attitude measurement

Researchers can choose from a myriad of scale formats, such as paired comparisons, graphical, semantic differentials, or Stapel scales to measure consumers' attitudes. An early distinction contrasts Thurstone's (1928) and Likert's (1932) approaches to attitude scaling. In Thurstone's terminology, an item refers to a verbal qualifier or the verbal description of a scale anchor, such as "excellent" or "mediocre." Thurstone's (1927) law of comparative judgment entails a general theoretical model that uses empirical data about item comparisons to develop measurement scales. Thurstonian approaches to scale development include equal appearing intervals, successive intervals, and paired comparisons (Thurstone, 1954). All the methods require the scale developers to select a large number of appropriate items that span the entire attitudinal continuum they wish to measure. Respondents then evaluate the candidate items to indicate the item position on this continuum. The procedures continue to the calibration phase, which involves the choice of an appropriate subset of items as verbal scale anchors, according to their psycho-numerical characteristics. Figure 1 depicts calibrated quality scales for the USA (Myers and Warner, 1968) and France (Angelmar and Pras, 1978). Our study extends this literature with a Japanese verbal rating scale for quality measurement.

Likert (1932) critiqued Thurstonian scale construction methods as laborious and promised simpler development processes. On a Likert scale, respondents indicate their degree of agreement with positively or negatively worded items (e.g. Kwok *et al.*, 2006). Therefore, the term "item" refers to the statements or the survey questions (Figure 2).

Researchers can use the data from sets of highly correlated items to build summative indicators that estimate a respondent's position on the underlying construct or to form latent variables that indicate the construct relations (Chung *et al.*, 2008; Nguyen and Barrett, 2006). Likert never provided theoretical justifications for this multi-item construct indicator approach, but rationales emerged under the label of test theory. Classical test theory involves a diverse body of psychometric methods that can identify the most discriminating, converging, and reliable items according to discrimination indexes, item-to-total correlations/-deletion coefficients, or loadings from factor analysis (e.g. Churchill, 1979; Witkowski *et al.*, 2003). The final multi-item Likert scale consists of a reasonably small set of items that appear optimal with regard to some criteria, such as a Cronbach's α greater than 0.7, a widely accepted condition for measurement reliability.

2.1 Advantages of the Likert approach

Likert-type scales are by far the most common survey instrument for attitude measurement; at least three reasons account for this great popularity: conformity with current research practice, ease of scale construction, and standards for measurement evaluation that align with test theory.

2.1.1 Conformity with current research practices. Scientific progress results mostly from researchers building on previous researchers' work. Researchers who want to test new or alternative hypotheses generally build on existing construct measurements, which helps strengthen the contributions that the new study makes to the existing literature. Therefore, the dominance of the Likert format may relate to the ever-growing number of studies that borrow multi-item measurement scales and standards from previous studies, which themselves use the Likert approach.

2.1.2 Ease of interval scale construction. Interval-level data offer good measurement precision and enable survey researchers to perform almost any mathematical operation they wish. An interval scale contains units of measurement that respondents perceive as equidistant. Likert scales appear to deliver a defensible approximation of interval-level

U.S. version (Myers & Warner, 1968)	<i>To me, brand X product quality is ...</i>						
	Horrible	Bad	Moderately Poor	Neutral	Pleasant	Delightful	Fantastic
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
French version (Angelmar & Pras, 1978)	<i>D'après moi, la qualité de la marque X est ...</i>						
	Terrible- mauvaise	Remarqu- ablement faible	Assez faible	Moyenne	Très correcte	Superbe	Extra- ordinaire
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Figure 1. Calibrated verbal rating scales to measure quality

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I received excellent service.	<input type="checkbox"/>				
I received service better than other firms' standards.	<input type="checkbox"/>				
I received superior service.	<input type="checkbox"/>				
I received high-quality service.	<input type="checkbox"/>				

Source: Iglesias (2004)

Figure 2. Multi-item Likert scale of customer perceived quality

data, whereas verbal rating scales evidently do not. For example, Westbrook (1980) and Chee and Peng (1996) use a “delighted” to “terrible” (D-T) verbal rating scale from sociology to measure customers’ quality perceptions[1], but they fail to offer evidence to support the equidistance assumption underlying the D-T scale’s verbal anchors. The interpretability of empirical data becomes limited when no such evidence is provided. Because Thurstone scales require a preliminary calibration phase to ensure the equivalence of interval-level data, whereas Likert-type scales can avoid this step, researchers may consider the Likert format more convenient.

2.1.3 Test theory. According to Churchill (1979), multi-item indices enjoy an advantage because single-item specificities get averaged out, relatively fine distinctions among respondents can be made, and measurement reliability tends to increase with the number of items in a combination. Likert scales are convenient for batteries of survey questions and facilitate multi-item measurement (see the example in Figure 2). In contrast, Thurstone scales are restricted by the availability of appropriate verbal expressions. Imagine, for example, developing additional verbal rating scales for Figure 1; the number of items with (exactly) the same metrics is limited in both languages. Some common interpretations of test theory that require empirical steps to evaluate the measurement (e.g. Anderson and Gerbing, 1988) cannot be performed with verbal rating scales, because the same procedures cannot apply to the multi-item scales. For example, researchers cannot compute Cronbach’s α , Jöreskog’s ρ , interitem correlations, and average variance extracted for single questions, and confirmatory factor analysis is inappropriate because it requires three indicators to justify or identify a latent construct (e.g. Anderson and Gerbing, 1988; Bagozzi, 1977, 1981; Fornell and Larcker, 1981; Jöreskog and Sörbom, 1982). Thurstone scales require different standards for the measurement evaluation, and widely accepted procedures to evaluate them do not exist.

2.2 Advantages of the Thurstone approach

The preceding arguments may prompt researchers to favor Likert-type measurement and overlook the advantages of Thurstone scales. However, three contrasting reasons may help them consider calibrated verbal rating scales as viable alternatives for attitude measurement, namely, the ease of operation, data symmetry, and the empirical response process.

2.2.1 Ease of operation. Single-item verbal rating scales (compared with multi-item Likert scales) reduce the length of questionnaires, which may be more cost efficient and improve respondents’ readiness to participate. A verbal rating scale based on colloquial, familiar, or current language also should facilitate performance of stimulus evaluations (e.g. Moxey and Sanford, 1991; Rohrmann, 2003). On a Likert scale, respondents must not only state a degree of agreement with each item but also cognitively interpret the meaning of each item as a representation of the same construct. This double evaluation may induce a potential source of measurement error, which can be avoided by Thurstone scales that describe the degrees of attitudinal positions directly. Multi-item indices commonly attempt to average out the single-item specificities and capture various aspects of a construct (Churchill, 1979). However, these different items may be located at different positions of the attitudinal continuum (e.g. “excellent service” vs “high-quality service”; Figure 2), in which case the empirical response process suffers, as we discuss subsequently.

2.2.2 Data symmetry. According to Westbrook (1980), satisfaction data frequently are skewed because the measurement is insufficiently sensitive to detect gradations of consumers’ sentiments. The D-T verbal rating scale can avoid skewness problems,

because it improves construct representation through explicit verbal reference to the gradation of affective meanings, which in turn improves the differentiation of responses at the upper end of the satisfaction continuum and reduces response bias (Westbrook, 1980). Data skewness also may stem from respondents' fatigue from answering long questionnaires with redundant questions (e.g. Reicheld, 1996). Arguably, respondents perceive multi-item scales as somewhat redundant. Thus, the benefit of greater reliability in the attitude scores of multi-item measures (Churchill, 1979) also entails a downside of increasing respondents' fatigue, which could result in skewed data. Such skewness may increase the reliability of the measurement, but not its validity.

2.2.3 Empirical response process. Thurstone (1927, 1928) proposed that the likelihood a person endorses an item depends on the distance between the item and the person's stand, which is best reflected by an ideal-point model. The ideal-point model is very different from the assumption of the dominance- (or cumulative-) response model on which Likert scale development procedures essentially rely (Coombs, 1964). In a dominance response model, a person endorses an item to the extent that his or her attitude position locates above that item (Andrich, 1996; Green, 1954; Roberts *et al.*, 1999). However, evidence from recent psychometric literature suggests that survey participants generally use some kind of ideal-point response when endorsing attitude statements (Andrich, 1996; van Schuur and Kiers, 1994). Roberts *et al.* (1999) and Stark *et al.* (2006) demonstrate that the item response functions (IRF) from Likert scales are more consistent with an ideal-point model than with a dominance model. The ideal-point model offers benefits for attitude scaling, because it can fit a monotonically increasing IRF of the dominance model properly but does not require this property. The dominance model therefore represents a special case of an ideal-point model. If the empirical response process generally follows an ideal-point model, it is reasonable to assume some psychometric advantages of Thurstone scales.

The differences in the empirical response process between Likert and Thurstone scaling approaches appear most vividly in the comparisons of the outcome of a moderate positive item (e.g. "good quality," "fine quality") and an extreme positive item (e.g. "superior quality," "excellent quality") for both ideal-point and dominance assumptions. As Figure 3 reveals, both response models make similar predictions about how people with moderate attitude positions respond to a moderately positive item, but increasing differences emerge for persons with extreme positive attitude positions. The ideal-point model suggests that respondents with extreme positive positions exhibit less agreement with a moderately positive item, because it does not reflect their attitude position well enough. In contrast, in the dominance model, people with extreme positions endorse a moderate positive item more than do those with moderate positions. Consequently, according to Roberts *et al.* (1999, p. 220), when non-monotonicity occurs in the response process, "Likert scores will suggest incorrectly that individuals with the most extreme Thurstone scores actually have more moderate opinions." This type of measurement error may be avoided when Likert scales use items that reflect extreme positions; however, Likert scales with moderate items are more common in current survey research (see Table I).

Furthermore, validity problems due to inappropriate, moderate Likert items appear particularly likely when respondents get ranked according to their trait position. Personality assessment studies show that a misspecification of the response process can lead to biased ranking results, with two main consequences: partially invalid behavior predictions and reduced utility of personality measures for employment and career decisions (Roberts *et al.*, 1999; Stark *et al.*, 2006). Similar measurement errors

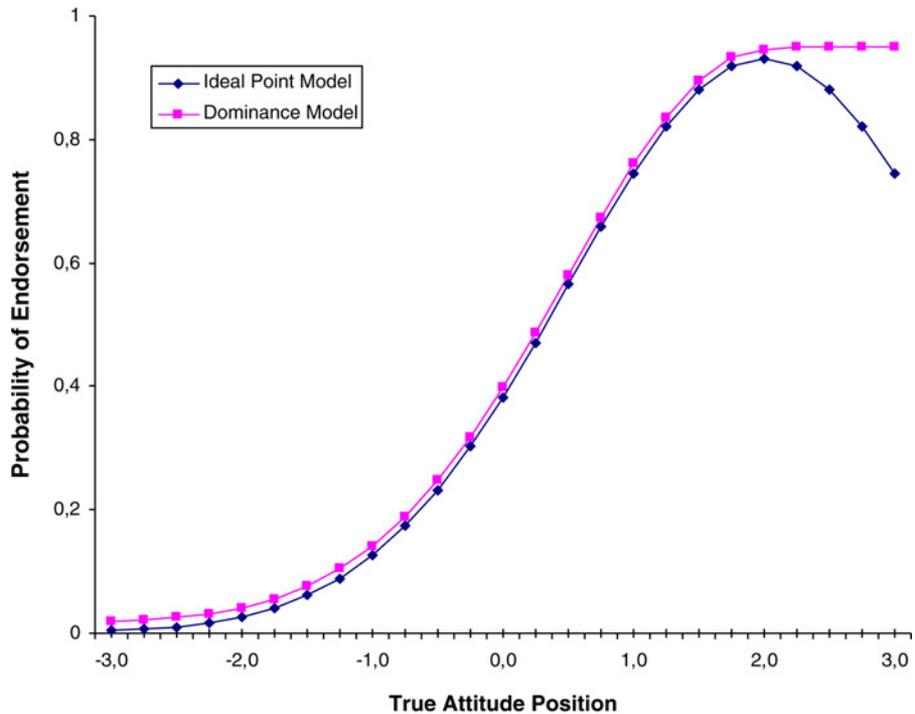


Figure 3.
Item response functions
for moderate positive
item, dominance- and
ideal-point assumptions

Source: Roberts *et al.* (1999) and Stark *et al.* (2006)

may occur when consumers provide product quality assessments with inappropriate Likert scales. For example, marketers may wish to derive consumers' price acceptance from their product quality evaluations, yet the identification of those buyers with the highest price acceptance will be biased when they are rank-ordered according to a Likert measurement with moderate items.

3. Verbal rating scale development

In this section, we outline the three steps in a Thurstone scale development procedure: identification of appropriate items, scale calibration, and measurement evaluation (reliability and validity testing). We use data collected from student samples, which are appropriate for this illustrative purpose (e.g. Wong *et al.*, 2008).

3.1 Item identification

In the first step, we must identify a set of verbal qualifiers that describe different levels of product quality. We asked 22 Japanese consumers to write down some expressions they would use in daily life to describe product quality. During thoughtful briefings and to facilitate their task, we told the respondents to imagine a discussion with friends about the quality of mobile phones (Figure 4). They produced a list of 45 different verbal qualifiers, which represent the candidate items we use for the scale development. In addition, we adopted seven candidate items from Myers and Warner (1968) and Angelmar and Pras (1978).

Authors	Construct	Examples
Burton <i>et al.</i> (1998)	Attitude toward private-label products	<ul style="list-style-type: none"> • Buying private label brands makes me feel good • For most product categories, the best buy is usually the private label brand • When I buy a private label brand, I always feel that I am getting a good deal
Newell and Goldsmith (2001)	Perceived corporate credibility	<ul style="list-style-type: none"> • XYZ corporation has great amount of experience • XYZ corporation is skilled in what they do • XYZ corporation does not have much expertise
Lages <i>et al.</i> (2005)	Relationship quality in export markets	<ul style="list-style-type: none"> • This importer frequently discussed strategic issues with us • This importer rarely talked with us about its business strategy • The parties involved had continuous interaction during implementation of the strategy
Ewing and Napoli (2005)	Nonprofit brand orientation	<ul style="list-style-type: none"> • Invest adequate resources in product/service improvements that provide better value to stakeholders • Create a brand/sub-brand that is well thought and understood by our staff • Develop a good understanding of the images/associations that our stakeholders make with the brand
Bearden <i>et al.</i> (2006)	Long-term orientation	<ul style="list-style-type: none"> • Respect for tradition is important to me • I plan for the long term • I value a strong link to my past

Table I. Moderate positive items in multi-item Likert scales

Suppose you are talking with friends about the quality of different brands and models of mobile phones. To express your opinion, you might state that the quality of your favorite mobile phone model or brand is “*really good*” or “*superior*,” but you might call the quality of another model or brand just “*acceptable*” and yet another “*quite poor*.”

The aim of this study is to collect expressions you may use to express your opinion about product quality. Please provide such expressions by completing the sentences below. We are interested in collecting as many different expressions as possible. So even if two expressions that come to your mind seem to be very similar, please note them anyway. Just make sure that the expressions describe degrees of product quality in a meaningful way.

<i>Please complete</i>	
The quality of model/brand X mobile phone is
The quality of model/brand X mobile phone is
The quality of model/brand X mobile phone is

Figure 4. Identification of verbal expressions to describe product quality

3.2 Verbal rating scale calibration

The aim of the scale calibration process was to select, from among the 52 candidate items, an appropriate subset that would fit the interval-scale measurement well; we outline this step next. To reduce sources of measurement error, we cross-checked the candidate items with four native Japanese speakers, who confirmed all items were meaningful and applicable to product quality measurements. Next, we employed Thurstone’s method to determine equally appearing intervals: We asked a sample of

104 Japanese consumers to sort the 52 candidate items into 11 successive numerical categories that ranged hierarchically from (−5) “The worst I could say about quality” to (+5) “The best I could say about quality.” This step may be confusing to survey respondents, who commonly perform stimulus evaluations (e.g. evaluations of brands or products), but not for the response evaluations requested in our study. To avoid misunderstanding and facilitate their task, we thoroughly briefed respondents and gave them careful instructions, together with the scenario in Figure 5.

According to Thurstone, the selected items should evenly span the entire attitudinal continuum. Item means or medians provide numerical representations of their position on this continuum, and standard deviations of the means indicate the amount of item ambiguity. The smaller this standard deviation is, the more consistently the item is rated by the respondents. In our study, we selected items for which the mean distances were as equal as possible, and we discarded those items with large standard deviations. Therefore, we obtain a seven-anchor verbal rating scale with almost equidistant items and low item ambiguity (Table II). Moreover, the items possess close to normal distributions, as indicated by their low and insignificant skewness and kurtosis results. We therefore may assume that the new scale is relatively homogeneously interpreted, and it is appropriate to generate interval-level data from customer surveys. By asking Japanese consumers to evaluate product quality by selecting the one scale item that best reflects their opinion, researchers can investigate mean differences across groups of respondents or perform other statistical analyses for which interval-level data are required.

3.3 Scale validity

In study settings with multi-item construct indicators, the evaluation of measurement validity commonly relies on empirical interpretations of test theories, which are not applicable to single-item Thurstone scales. However, the evaluations of the validity of these scales still can use test theories, namely, measurement reliability and construct validity. A measure is reliable when it provides the same results over and over again, as long as that which is being measured does not change (Trochim, 2000). To test the reliability of Thurstone scales, researchers can use test–retest strategies or split-half techniques and subsample comparisons of the measured scores.

Suppose you are talking with friends about the quality of different brands and models of mobile phones. You might state that the quality of your favorite model or brand is “*really good*” or “*superior*,” but you might call the quality of another model or brand just “*acceptable*” and yet another “*quite poor*.”

The aim of this study is to understand what you really mean when you employ such expressions of quality. Please tell us, for each of the following 52 expressions, whether you think they express *the worst thing* or *the best thing you could say about quality*, or *any level in between these extremes*. Remember that you are not evaluating a specific model or brand but rather each expression’s meaning.

	This is the worst thing I could say about quality					This is the best thing I could say about quality					
	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Candidate item 1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Candidate item 2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Candidate item 52	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Figure 5.
Response scaling

English	Japanese	Japanese script	Std. dev.	Skewness (z)	Kurtosis (z)	Median (Δ)	Mean (Δ)
Worst	Saitei	最低	0.57	1.86 (7.75)*	2.47 (5.15)*	1	1.29
Bad	Warui	悪い	0.83	-0.38 (-1.60)	-0.25 (-0.53)	3 (2.00)	2.77 (1.48)
A little bad	Sukoshi warui	少し悪い	0.58	0.07 (0.30)	-0.30 (-0.62)	4 (1.00)	4.34 (1.57)
Neutral	Dochiratomo ienai	どちらとも言えない	0.53	-0.29 (-1.21)	0.48 (1.01)	6 (2.00)	5.88 (1.55)
So so	Soko soko	そこそこ	1.04	-0.32 (-1.33)	-0.33 (-0.69)	7 (1.00)	7.28 (1.39)
Satisfying	Mannzoku dekuru	満足できる	1.08	-0.04 (-0.16)	0.09 (0.18)	9 (2.00)	8.68 (1.40)
Very good	Totemo yoi	とても良い	0.74	-0.85 (-3.56)*	0.71 (1.48)	10 (1.00)	10.24 (1.56)

Note: *Significant at the 0.01 level
Source: Hair *et al.* (2006)

Table II.
 Psychometric properties
 of the seven-item scale

Reliability is a necessary but not sufficient condition for measurement validity. A measure attains construct validity when its operationalization accurately reflects the construct that it attempts to measure. Cronbach and Meehl's (1955) view of construct validity, in terms of nomological validity, offers a grounded theory approach to validity testing. According to these authors, linkages between measured constructs should be specified within a theoretically affirmed nomological network. A new measure therefore has nomological validity if these linkages hold empirically.

We investigate the nomological validity of our new verbal rating scale by studying how consumers' ethnocentrism affects their quality perceptions of branded products in different competitive situations. Previous research provides significant evidence that consumers' quality perceptions depend on their ethnocentrism and the brands' country of origin (e.g. Aron and Kenny, 2002; Hamin, 2006; Lundstrom *et al.*, 1998; Sohail, 2005; Wong *et al.*, 2008; Yamen, 2008). Shimp and Sharma (1987) describe ethnocentrism as the beliefs held by consumers about the appropriateness and morality of purchasing imported products. According to these authors, highly ethnocentric consumers tend to believe that purchasing foreign products is wrong because it hurts the domestic economy, causes the loss of jobs, and is plainly unpatriotic. Foreign products become objects of contempt to highly ethnocentric consumers, who make quality inferences on the basis of their perception of the country of origin of the brand; products from domestic brands yield higher brand quality evaluations than do those produced by foreign brands. Nonethnocentric consumers, in contrast, tend to appreciate foreign products on their own merits and without consideration for where they are made.

To evaluate consumers' ethnocentric tendencies, Shimp and Sharma (1987) propose the CETSCALE, a multi-item measurement instrument frequently employed to predict consumers' buying behavior (e.g. Netemeyer *et al.*, 1991). However, variations in the predictive validity of the CETSCALE exist, depending on contextual factors such as the foreign products' quality and the availability of domestic product alternatives in a category. Studies by Supphellen and Rittenburg (2001) and Yoo and Donthu (2005) also

indicate that high consumer ethnocentrism (CE) does not necessarily lead to negative perceptions of foreign brands when those brands are commonly regarded as offering higher quality than do domestic brands. For example, Kwok *et al.* (2006) state that Chinese consumers often perceive foreign brands as having higher quality. Similarly, when foreign brands dominate a product category (i.e. when acceptable domestic alternatives are not available in a product category), even highly ethnocentric consumers may accept the purchase of foreign brands and recognize their high quality. As Nijssen and Douglas (2004) show, Dutch consumers, despite their expressions of animosity toward Germany, tend to evaluate German cars favorably because they cannot find domestic alternatives. However, when domestic brands of the product categories are available, feelings of animosity overwhelm positive perceptions of foreign product quality, so ethnocentric consumers appear reluctant to buy foreign products. That is, feelings of ethnocentrism come to the fore only when a domestic brand is available. We build on these conceptualizations and evidence from previous literature to formulate two validity assumptions, which we test empirically to determine the nomological validity of the new verbal rating scale.

Validity Assumption 1. When neither domestic nor foreign brands clearly dominates a product category, consumers with high ethnocentrism ratings make quality inferences on the basis of the perceived country of origin of the brand, such that products from domestic brands, compared with products from foreign brands, yield higher quality evaluations. On the contrary, consumers with low ethnocentrism make such quality inferences to a much lesser degree.

Validity Assumption 2. When foreign brands dominate a product category, consumers' quality evaluations of foreign brands are independent of whether consumers are high or low on ethnocentrism.

3.3.1 Data collection and study design. We conducted a study with 151 Japanese consumers who complete the new verbal rating scale (Figure 5) to evaluate the product quality of 32 well-known brands in six product categories. Two versions of the questionnaire present the brands in different orders.

The six product categories are all relevant to our sample of respondents. For some categories, both domestic and foreign brand alternatives are available on the Japanese market, whereas for others, the product category is dominated by foreign brands (Table III). Specifically, we include cars, laptop computers, and soft drinks as categories

Product category	Brand origin		Competitive situation
	Japanese	Foreign	
Cars	Daihatsu; Honda; Mazda; Mitsubishi; Nissan; Toyota	BMW; Ford; Hyundai; Mercedes-Benz; Volkswagen; Peugeot	Domestic and foreign brand alternatives available
Laptop computer	Fujitsu; Panasonic; Sony	Apple; Dell; Hewlett Packard	
Soft drinks	Asahi; Itoen; Kirin; Suntory	Coca-Cola; Nescafe; Pepsi Cola; Starbucks	Foreign brands dominant
Internet search engines	–	Google; Yahoo!	
Sport shoes	–	Adidas; Nike	
Amusement parks	–	Disney; Universal Studios Japan	

Table III.
Product categories and Japanese domestic and foreign brands

in which Japanese brands compete heavily with foreign brands. In contrast, Internet search engines (Yahoo, Google), sport shoes (Adidas, Nike), and amusement parks (Disney, Universal Studios) are dominated by foreign brands in the Japanese market. Finally, to measure respondents' degree of ethnocentrism, we offered the 10-item CETSCALE (Shimp and Sharma, 1987), after applying a careful translation-back translation procedure with two bilingual (English/Japanese) speakers. The respondents indicated their stance toward each CETSCALE item on five-point Likert scales.

3.3.2 Measurement evaluation. Before testing the nomological validity of the new verbal rating scale, we investigate its reliability, as well as the dimensionality and reliability of the CETSCALE. We find no statistically significant differences in the brand quality evaluations according to the brand rotation ($p > 0.05$), which provides initial evidence of the reliability of the new verbal rating scale. We also randomly select half of the respondents and compare brand quality ratings across the two subgroups. Again, we determine no statistical differences ($p > 0.05$), in support of the split-half reliability of the new verbal rating scale. Next, the results from an exploratory factor analysis of the ten-item CETSCALE suggest that the anticipated one-dimensional structure of this scale holds after we exclude two items with weak loadings on the first factor. The remaining eight items explain 71.7 per cent of the variance, and the Cronbach's alpha is high ($\alpha = 0.94$). These results confirm Netemeyer *et al.*'s (1991, p. 324) reports of a one-dimensional CETSCALE structure in Japan with internal consistency reliabilities as high as $\alpha = 0.91$.

3.3.3 Nomological validity of the new verbal rating scale. We used a two-way, between-subjects factorial design to test the two validity assumptions. Respondents may perform interdependent brand evaluations, especially of brands within the same product category; for example, a consumer could prefer Coca-Cola and for that reason alone reject Pepsi. To avoid statistical problems in the dependent measures, we aggregate the brand evaluations within the treatment conditions we outline subsequently (Hair *et al.*, 2006, p. 409). That is, we assume that consumers evaluate the quality of brands independently across the categories, such that they perceive cars relatively independently of laptop computers or soft drinks. Factor 1 (*competitive situation*) consists of three levels that we controlled through the choice of product categories and brands (Table II). The three levels are

- (1) domestic brands/domestic and foreign brand alternatives available;
- (2) foreign brands/domestic and foreign brand alternatives available; and
- (3) foreign brands/foreign brands dominant.

Our Factor 2 (CE) is an individual characteristic that we observe from the data. We use a median split, based on the eight CE indicators, to differentiate between high and low ethnocentric respondents.

Levene tests suggest variance homogeneity within the product categories for 28 of the 32 brands, but variance homogeneity is not supported across the six treatment conditions. This violation of an analysis of variance (ANOVA) assumption should have little effect on the reliability of our results from group comparisons though, because the groups we compare are of approximately equal size (Hair *et al.*, 2006). The effect of the interaction between CE and competitive situation is highly significant ($p < 0.001$; $F(41.419, 5)$), as we show in Table IV. When domestic brand alternatives are available (cars, computer, soft drinks), high CE consumers evaluate foreign brands significantly less favorably ($\bar{x} = 4.51$) than domestic brands ($\bar{x} = 5.25$), as well as significantly

Table IV.
Between-group
comparisons

Competitive situation (<i>I</i>)	\bar{x} ;	Consumer ethnocentrism (<i>I</i>)						Post-hoc comparisons		
		Low (1)			High (2)			Test	Δ	Scheffe sig.
Domestic brands/domestic and foreign brands available (1)	5.14	1.39	1066	5.25	1.21	897	$I_1J_1-I_1J_2$ $I_1J_2-I_2J_2$	-0.11 0.74	0.728 0.000	
Foreign brands/domestic and foreign brands available (2)	5.02	1.60		4.51	1.75		$I_2J_1-I_1J_1$ $I_2J_2-I_2J_1$	-0.12 -0.51	0.580 0.000	
Foreign brands/foreign brands dominant (3)	5.44	1.55	492	5.48	1.33	414	$I_3J_1-I_3J_2$	-0.04	0.999	

less favorably than do consumers with low CE (\bar{x} ; = 5.02). Moreover, the low CE consumers exhibit similar evaluations for both domestic (\bar{x} ; = 5.14) and foreign (\bar{x} ; = 5.02) brands, which confirms validity assumption 1.

When foreign brands have a dominant market position (search engines, sport shoes, theme parks), no such effect of CE exists. That is, consumers with low and high CE provide similar quality evaluations of foreign brands (CE_{low} \bar{x} ; = 5.44; CE_{high} \bar{x} ; = 5.48), in line with validity assumption 2. We confirm these interpretations of the mean differences with post-hoc Scheffe tests (Table III). Overall, our results support both validity assumptions; we thus conclude that the new verbal rating scale achieves nomological validity.

4. Conclusions

Perhaps the most important conclusion from this article is that both theoretical and empirical evidence exists to consider Thurstone or verbal rating scales as at least viable alternatives to the more frequently employed Likert approach to attitude scaling. To shed more light on the use of verbal rating scales, we illustrate an actual Thurstone scale development procedure, including the item generation, scale calibration, and measurement evaluation steps, and test for its reliability and validity. Attitude measurement remains an important concern for all management fields, yet verbal rating scales are rarely employed, nor do standardized approaches to their development and evaluation exist. To this end, we offer theoretical and practical guidance, as well as sources of inspiration, for researchers who might consider employing or developing verbal rating scales.

Our new verbal rating scale also provides several practical advantages for measuring perceived product quality in Japan. First, it offers a ready-to-apply measurement instrument for many types of data collection, including paper-and-pencil interviews, online questionnaires, and telephone surveys. The scale calibration process ensures that researchers may gather interval-level data with this scale, which offers greater flexibility for empirical analyses. Second, using a single question for stimuli evaluation, rather than multiple items, reduces the length of questionnaires. Therefore, the new scale should be more cost effective than multi-item quality measurements and reduce respondents' fatigue during the evaluation task. Third, respondents may find it easier to perform their quality evaluations on the new scale, because it uses common language expressions as scale anchors. Survey participants also may perceive a single, verbal rating scale as less cognitively demanding than multi-item Likert scales. Short

questionnaires and comprehensible measurement help increase respondents' willingness to share information and may reduce the likelihood of misunderstandings. We provide the final version of our proposed verbal rating scale in Figure 6.

Beyond these practical advantages, verbal rating scales offer appropriate operationalizations of the empirical response process, with its implications for measurement validity. Previous research reveals that misspecifications in the response process can lead to biased results due to the ranking of individual respondents (Roberts *et al.*, 1999; Stark *et al.*, 2006). According to Stark *et al.* (2006, p. 37), "the inclusion of just a few items that do not meet the assumptions of a dominance model can markedly change the rank order of high-scoring individuals," which then leads to invalid behavioral predictions and reduced measurement utility. However, examples from recent literature suggest that studies frequently employ Likert scaling procedures even when the assumptions of the dominance response process cannot be met because the item wording includes moderate rather than extreme terminology (Appendix). Such a misspecification of the response process may have little influence on the accuracy of the predictions of construct relations though. The strength of the causal relations, such that higher quality leads to higher price acceptance, may be predicted sufficiently well by a poorly specified dominance approach with moderate Likert items, because correlation coefficients are relatively insensitive to rank-order changes. More research is needed to explore in greater detail why and in which conditions the dominance or ideal-point approach to item responses is more appropriate for improving measurement validity.

We also recognize some limitations of our study that could benefit from additional research. First, the evidence that the new scale delivers interval-level data should be attenuated. Interval data can be assumed only to the extent that the respondents to our scale calibration study perceive the intervals between numerical anchors as truly equidistant (Figure 5). We do not test this assumption. Furthermore, response bias may exist in the calibration study, because the meaning of each candidate item may not be reflected accurately by the numerical categories. For example, respondents may associate good or bad luck with some numbers, such that their numerical preferences influence their response behavior. Studies of response bias in verbal rating scale calibration procedures would provide meaningful insights into an underinvestigated field.

Second, we claim construct validity for our new scale, but we also note some constraints on that claim. Our study of construct validity is limited to how perceived product quality may be influenced by CE and the competitive situation of domestic and foreign brands. Further research should broaden this demonstration of construct validity, perhaps by building different nomological networks and relating the new measurement to other constructs, such as customer satisfaction, trust, company reputation, or employee satisfaction, as well as considering the potential moderators or mediators of such relations. A vast body of literature pertaining to such relations exists, but empirical investigations mostly refer to Likert-type measurements. Are these results replicable when we use Thurstone scales, and why might differences or communalities emerge? We also recognize that the reasoning we adopted to

To me, brand X product quality is ...						
最低 (worst)	悪い (bad)	少し悪い (a little bad)	どちらとも 言えない (neutral)	そこそこ (so so)	満足できる (satisfying)	とても良い (very good)
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Figure 6. Calibrated verbal rating scale

demonstrate the nomological validity of the new scale conflicts slightly with the reasoning we adopted to choose Thurstone scales for our attitude measurement. That is, we used a Likert-scaling procedure to operationalize the CE construct, but we had previously argued that Likert-scaling procedures may lead to validity problems. Because the CETSCALE items in our study express extreme attitude positions (rather than moderate positions), we believe the dominance response process should work well for this scale and do not believe that the misspecification of the response process presents a significant threat to our assessment of scale validity.

Third, we note the limited generalizability of the new scale; it relies on data from samples of respondents who are obviously not representative of the wider Japanese consumer population. Our consumer samples consist of management students from a major Japanese university, who in 2007 lived in the Osaka region, which means they may use and perceive verbal expression differently than would consumers from other lifestyle segments or geographical regions. Further research should investigate if and to what extent the new scale generalizes across consumer segments. Scale developers should recognize that language evolution is a dynamic process, so frequently employed expressions at one point in time may disappear, even as new language standards evolve. Kawashima (2001) highlights the importance and effects of changes in the Japanese language in particular. Later researchers inevitably must address the appropriateness of earlier measures; some of the expressions Myers and Warner (1968) and Angelmar and Pras (1978) used in their time likely are no longer appropriate in verbal rating scales for quality measurement today.

Another interesting area for future research may be to develop verbal rating scales for cross-cultural comparisons. For example, Wongtada and Rice (2008) use item response theory to reveal that interpretations of measurement scores by Thai and Egyptian employees require adjustments to be meaningful. Instead of conducting a posteriori adjustments of cross-cultural survey data or measurement instruments, researchers might try to calibrate a priori culturally or language-invariant verbal rating scales, especially in regions that contain significant language diversity. For example, Chinese is a macro-language that comprises 13 sub-languages (ISO 639-3). Do verbal rating scales written in Chinese script require different expressions to improve the validity of comparisons across Chinese consumers who speak different languages or dialects?

Finally, though the development of a verbal rating scale was the primary objective of this study, we also note its contributions to CE literature. Specifically, findings by Nijssen and Douglas (2004), which were limited to a single, foreign brand-dominated product category (cars) in the Netherlands, receive confirmation from our study, which features a culturally different context and five additional product categories (in addition to cars, laptop computers, soft drinks, sport shoes, theme parks, and internet search engines).

Note

1. The D-T scale consists of seven items: delighted, pleased, mostly satisfied, mixed (about equally satisfied and dissatisfied), mostly dissatisfied, unhappy, and terrible.

References

- Anderson, J.C. and Gerbing, D.W. (1988), "Structural equation modeling in practice: a review and recommended two-step approach", *Psychological Bulletin*, Vol. 103, pp. 411-23.

-
- Andrich, D. (1996), "A hyperbolic cosine latent trait model for unfolding polytomous responses: reconciling Thurstone and Likert methodologies", *British Journal of Mathematical and Statistical Psychology*, Vol. 49, pp. 347-65.
- Angelmar, R. and Pras, B. (1978), "Verbal rating scales for multinational research", *European Research*, Vol. 6, pp. 62-6.
- Aron, O.C. and Kenny, L. (2002), "Understanding the younger Singaporean consumers' views of Western and Eastern brands", *Asia Pacific Journal of Marketing and Logistics*, Vol. 14, pp. 54-79.
- Bagozzi, R.P. (1977), "Convergent and discriminant validity by analysis of covariance structures: the case of the affective, behavioral, and cognitive components of attitude", *Advances in Consumer Research*, Vol. 4, pp. 11-18.
- Bagozzi, R.P. (1981), "Evaluating structural equation models with unobservable variables and measurement error: a comment", *Journal of Marketing Research (JMR)*, Vol. 18, pp. 375-81.
- Bearden, W.O., Money, R.B. and Nevin, J.L. (2006), "A measure of long-term orientation: development and validation", *Journal of the Academy of Marketing Science*, Vol. 34, pp. 456-67.
- Burton, S., Lichtenstein, D.R., Netemeyer, R.G. and Garretson, J.A. (1998), "A scale for measuring attitude toward private label products and an examination of its psychological and behavioral correlates", *Journal of the Academy of Marketing Science*, Vol. 26, pp. 293-306.
- Buzzell, R.D. and Gale, B.T. (1987), *The PIMS Principles: Linking Strategy to Performance*, The Free Press, New York, NY.
- Chee, L.K. and Peng, N.K. (1996), "Customer orientation and buyer satisfaction: the Malaysian housing market", *Asia Pacific Journal of Management*, Vol. 13, pp. 101-16.
- Chung, J.-E., Sternquist, B. and Chen, Z. (2008), "Japanese retail-buyer-supplier relationships: does performance matter?", *Asia Pacific Journal of Marketing and Logistics*, Vol. 20, pp. 55-75.
- Churchill, G.A. (1979), "A paradigm for developing better measures of marketing constructs", *Journal of Marketing Research*, Vol. 16, pp. 64-73.
- Coombs, C.H. (1964), *A Theory of Data*, Wiley, New York, NY.
- Cronbach, L. and Meehl, P. (1955), "Construct validity in psychological tests", *Psychological Bulletin*, Vol. 52, pp. 281-302.
- Ewing, M.T. and Napoli, J. (2005), "Developing and validating a multidimensional nonprofit brand orientation scale", *Journal of Business Research*, Vol. 58, pp. 841-53.
- Fornell, C. and Larcker, D.F. (1981), "Evaluating structural equation models with unobservable variables and measurement error", *Journal of Marketing Research*, Vol. 18, pp. 39-50.
- Green, B.F. (1954), "Attitude measurement", in Lindzey, G. (Ed.), *Handbook of Social Psychology*, Addison-Wesley, Reading, MA, pp. 335-69.
- Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E. and Tatham, R.L. (2006), *Multivariate Data Analysis*, Pearson Prentice Hall, Upper Saddle River, NJ.
- Hamin, G.E. (2006), "A less-developed country perspective of consumer ethnocentrism and 'Country of origin' Effects: Indonesian evidence", *Asia Pacific Journal of Marketing and Logistics*, Vol. 18, pp. 79-92.
- Iglesias, V. (2004), "Preconceptions about service: how much do they influence quality evaluations?", *Journal of Service Research*, Vol. 7, pp. 90-103.
- Jöreskog, K.G. and Sörbom, D. (1982), "Recent developments in structural equation modeling", *Journal of Marketing Research (JMR)*, Vol. 19, pp. 404-16.
- Kawashima, N. (2001), *Language Planning and Language Change in Japan*, Curzon Press, Richmond.

- Kwok, S., Uncles, M. and Huang, Y. (2006), "Brand preferences and brand choices among urban Chinese consumers", *Asia Pacific Journal of Marketing and Logistics*, Vol. 18, pp. 163-72.
- Lages, C., Lages, C.R. and Lages, L.F. (2005), "The RELQUAL scale: a measure of relationship quality in export market ventures", *Journal of Business Research*, Vol. 58, pp. 1040-8.
- Likert, R. (1932), "A technique for the measurement of attitudes", *Archives of Psychology*, Vol. 140, pp. 5-53.
- Lundstrom, W.J., Lee, O.W. and White, S.D. (1998), "Factors influencing Taiwanese consumer preference for foreign-made white goods: USA versus Japan", *Asia Pacific Journal of Marketing and Logistics*, Vol. 10, pp. 5-29.
- Masaaki, I. (1986), *Kaizen: The Key to Japan's Competitive Success*, McGraw-Hill, New York, NY.
- Moxey, L.M. and Sanford, A.J. (1991), "Context effects and the communicative functions of quantifiers: implications for their use in attitude research", in Schwarz, N. and Sudman, S. (Eds), *Context Effects in Social and Psychological Research*, Springer, New York, NY, pp. 279-96.
- Myers, J.H. and Warner, W.G. (1968), "Semantic properties of selected evaluation adjectives", *Journal of Marketing Research*, Vol. 5, pp. 409-12.
- Netemeyer, R.G., Durvasula, S. and Lichtenstein, D.R. (1991), "A cross-national assessment of the reliability and validity of the CETSCALE", *Journal of Marketing Research*, Vol. 28, pp. 320-7.
- Newell, S.J. and Goldsmith, R.E. (2001), "The development of a scale to measure perceived corporate credibility", *Journal of Business Research*, Vol. 52, pp. 235-47.
- Nguyen, T.D. and Barrett, N.J. (2006), "The adoption of the internet by export firms in transitional markets", *Asia Pacific Journal of Marketing and Logistics*, Vol. 18, p. 29.
- Nijssen, E.J. and Douglas, S.P. (2004), "Examining the animosity model in a country with a high level of foreign trade", *International Journal of Research in Marketing*, Vol. 21, pp. 23-38.
- Pecht, M. and Boulton, W.R. (1995), "Quality assurance and reliability in the Japanese electronics industry", in Boulton, W.R. (Ed.), *Electronic Manufacturing and Packaging in Japan*, International Technology Research Institute, Baltimore, MD.
- Reicheld, F.F. (1996), "The satisfaction trap", *Harvard Business Review*, Vol. 74, pp. 58-9.
- Roberts, J.S., Laughlin, J.E. and Wedell, D.H. (1999), "Validity issues in the Likert and Thurstone approaches to attitude measurement", *Educational and Psychological Measurement*, Vol. 59, pp. 211-33.
- Rohrmann, B. (2003), "Verbal qualifiers for rating scales: sociolinguistic considerations and psychometric data", available at: www.rohrmannresearch.net (accessed 15 September 2008).
- Shimp, T.A. and Sharma, S. (1987), "Consumer ethnocentrism: construction and validation of the CETSCALE", *Journal of Marketing Research*, Vol. 24, pp. 280-9.
- Sohail, S.M. (2005), "Malaysian consumers' evaluation of products made in Germany: the country of origin effect", *Asia Pacific Journal of Marketing and Logistics*, Vol. 17, pp. 89-105.
- Stark, S., Chernyshenko, O.S., Drasgow, F. and Williams, B.A. (2006), "Examining assumptions about item responding in personality assessment: should ideal point methods be considered for scale development and scoring?", *Journal of Applied Psychology*, Vol. 91, pp. 25-39.
- Supphellen, M. and Rittenburg, T.L. (2001), "Consumer ethnocentrism when foreign products are better", *Psychology & Marketing*, Vol. 18, pp. 907-27.
- Thurstone, L.L. (1927), "Three psychophysical laws", *Psychological Review*, Vol. 34, pp. 424-32.
- Thurstone, L.L. (1928), "Attitudes can be measured", *The American Journal of Sociology*, Vol. 26, pp. 249-69.

-
- Thurstone, L.L. (1954), "The measurement of values", *Psychological Review*, Vol. 61, pp. 47-58.
- Trochim, W. (2000), *The Research Methods Knowledge Base*, Atomic Dog Publishing, Cincinnati, OH.
- Van Schuur, W.H. and Kiers, H.A.L. (1994), "Why factor analysis is often the incorrect model for analyzing bipolar concepts, and what model can be used instead", *Applied Psychological Measurement*, Vol. 18, pp. 97-110.
- Westbrook, R.A. (1980), "A rating scale of measuring product/service satisfaction", *Journal of Marketing*, Vol. 44, pp. 68-72.
- Witkowski, T.H., Ma, Y. and Zheng, D. (2003), "Cross-cultural influences on brand identity impressions: KFC in China and the United States", *Asia Pacific Journal of Marketing and Logistics*, Vol. 15, pp. 74-88.
- Wong, C.Y., Polonsky, M.J. and Garma, R. (2008), "The impact of consumer ethnocentrism and country of origin sub-components for high involvement products on young Chinese consumers' product assessments", *Asia Pacific Journal of Marketing and Logistics*, Vol. 20, pp. 455-78.
- Wongtada, N. and Rice, G. (2008), "Multidimensional latent traits of perceived organizational innovation: differences between Thai and Egyptian employees", *Asia Pacific Journal of Management*, Vol. 25, pp. 537-62.
- Yamen, K. (2008), "Country of origin, brand image perception, and brand image structure", *Asia Pacific Journal of Marketing and Logistics*, Vol. 20, pp. 139-55.
- Yoo, B. and Donthu, N. (2005), "The effect of personal cultural orientation on consumer ethnocentrism: evaluations and behaviors of US Consumers toward Japanese products", *Journal of International Consumer Marketing*, Vol. 18, pp. 7-44.

About the authors

Boris Bartikowski is Associate Professor of Marketing at Euromed Management. He has been invited visiting professor at Kansai University, Osaka. He holds a joint doctoral degree from the University Paul-Cézanne (IAE-CERGAM) and the University of Augsburg. His research centers on consumer behavior, with a focus on empirical measurement and cultural diversity. Boris Bartikowski is the corresponding author can be contacted at: boris.bartikowski@euromed-management.com

Katsuyuki Kamei is Professor of Strategy and Risk Management at Kansai University, Osaka. He holds a D.E.A. degree in Management from the University Paul-Cézanne (IAE-CERGAM) and a PhD in management from Osaka City University. He is doing research in the fields of risk information disclosure, risk management in small to medium enterprises, and the marketing of financial institutions.

Jean-Louis Chandon is Professor of Marketing at the Paul Cézanne University Aix-Marseille (IAE-CERGAM). He holds a French state doctoral degree and a PhD from Northwestern University. He conducts research in the fields of empirical measurement and consumer behavior on the internet.