

A Thurstone Scale to Measure Consumers Perceptions of Product Quality in Japan

| | |
|------------------------------|---|
| 著者 | Bartikowski Boris, Kamei Katsuyuki |
| journal or publication title | 情報研究 : 関西大学総合情報学部紀要 |
| volume | 31 |
| page range | 107-113 |
| year | 2009-07 |
| URL | http://hdl.handle.net/10112/1014 |

A Thurstone Scale to Measure Consumers Perceptions of Product Quality in Japan

Boris BARTIKOWSKI^{*1} Katsuyuki KAMEI^{*2}

Abstract

Product quality is traditionally a central issue in Japanese industry processes and has been shown to be one of the most important factors to ensure long term business success. Thus, the appropriate measurement of consumers' quality perceptions is an essential consideration too. The article discuss advantages of verbal rating scales and follows a Thurstone approach to develop a measurement scale researchers and managers may use to measure product quality as perceived by Japanese consumers.

Key Words : Quality measurement; Thurstone; Verbal Rating Scale;

The authors gratefully acknowledge the financial support of the Kansai University for awarding a research fellowship to realize this project.

^{*1} Euromed Marseille School of Management ^{*2} Kansai University Faculty of Informatics

1. Introduction

Measuring peoples' opinions or attitudes on questionnaires with measurement scales is current practice in Management. To facilitate survey respondents' evaluation task and improve the quality of attitudinal measurement, researchers may use rating scales with verbal descriptions of the scale anchors, so-called Thurstone- or verbal-rating scales. The objective of this paper to develop a verbal rating scale that managers and researchers may use to measure Japanese consumers' perceptions of product quality.

The field of quality perceptions was chosen since quality has been shown to be a major determinant of sustainable business success. A finding of the PIMS studies was that a firm's product or service quality, relative to that of competitors, is "the most important single factor affecting a business unit's performance" (Buzzell and Gale 1987). Moreover, quality and reliability are since long integral considerations to Japan's industry and production processes. Slogans by Sharp Electronics "Quality First in Heart and Mind," or by Fujitsu "Quality built-in, with cost and performance as prime consideration" illustrate this point. Kaoru Ishikawa, one of the fathers of Total Quality Control Management (TQC) in Japan, outlined by 1968 that "quality comes first, not short-term profits." His "cause-and-effect" diagram suggests that product quality is essentially affected by variation of four factors, namely materials, equipment, processes, and measurement (Ishikawa 1982). Finally, "customer-driven quality" is one of the seven core values of the famous "Japan Quality Award." Our paper draws on these considerations and offers a verbal rating scale to measure product quality as perceived by Japanese customers.

We first revise and discuss the Thurstone approach of attitude scale construction. Next, we run a scale development study comprised by two stages: verbal item selection and scale calibration. Practical applications of the scale, limitations of the study, and future research directions are described in the conclusion.

2. The Thurstone Approach of Attitude Scaling

Thurstone's law of comparative judgment presents a general theoretical model that allows developing scales for psychological measurement from empirical data on item comparisons (Thurstone 1927). Researchers first select a large number of appropriate

items that span the entire attitudinal continuum they wish to measure. These candidate items are subsequently evaluated by a sample of respondents with regard to their position on the continuum. Next, the procedure requests calibrating the scale, which refers to the process of choosing an appropriate subset of items or verbal scale anchors based on their numerical characteristics. The following sections present three steps of Thurstone scale development, namely identification of appropriate items to describe product quality and calibration of a verbal rating scale.

2.1 Item Identification

The first step of our study was to identify a set of verbal expressions that describe different levels of product quality. To complete this task, we asked a sample of 22 Japanese consumers to write expressions they may use in daily life to describe product quality. Respondents were thoroughly briefed by the researchers, and to facilitate their task they were told to imagine they were discussing with friends about the quality of mobile phone models and brands. All respondents were experienced mobile phone users since many years, so it could be expected they had experienced a continuum of good and bad mobile phone qualities, which they could verbally describe. The given task is illustrated in Figure-1.

Suppose you are discussing with friends about the quality of different brands and models of mobile phones. To express your opinion you may state that the quality of your favorite mobile phone model or brand is “*really good*” or “*superior*,” but you may call the quality of another model or brand being just “*acceptable*,” and yet another one as “*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. Please note 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 write them despite. Just make sure that the expressions you write 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 1 Identification of Verbal Expressions to Describe Product Quality

This procedure resulted in a list with 45 different verbal qualifiers, or candidate items we could use for subsequent scale development. An additional review of the literature helped identifying additional items so that our final list had 52 candidate items, as shown in Figure-2. It should be noted that some of the English expressions from the literature could not be unambiguously translated into Japanese, and inversely no unambiguous English counterpart could be found for some of the Japanese expressions. Thus, some of the English translations shown in Figure-2 remain approximate and may not appropriately reflect emic meaning of the original Japanese expressions.

| | | | | |
|--------------------------|----------------------------|----------------------------|---------------------------|-------------------------|
| 非常に悪い (very bad) | 不満足 (not satisfying) | たいしたことない (not so good) | なかなか (so so) | とてもいい (fairly good) |
| 極めて悪い (extremely bad) | 駄目 (useless) | やばい (terrific) | 良い (good) | とても良い (very good) |
| 最悪 (worst) | 悪い (bad) | どちらとも言えない (neutral) | いい (delightful) | 素晴らしい (wonderful) |
| 全然駄目 (zero) | 意味(が)ない (meaningless) | 悪くはない (not bad) | 好ましい (delightful) | かなり良い (very good) |
| 最低 (worst) | 満足できない (not satisfying) | 普通 (normal) | 満足できる (satisfying) | とても素晴らしい (very good) |
| まったく駄目 (zero) | あまりよくない (not so good) | まし (rather good) | 便利 (useful) | 非常に良い (very good) |
| 使い物にならない (not usable) | 良くない (not good) | 手ごろな (reasonable) | よくできて(いる) (quite good) | めっちゃいい (very good) |
| ひどく悪い (very bad) | やや悪い (rather bad) | まあまあ (moderately good) | 結構良い (rather good) | 最高 (excellent) |
| かなり悪い (very bad) | 少し悪い (little bad) | 大丈夫 (OK) | 優れている (superior) | |
| とても悪い (very bad) | いまいち (not very good) | そこそこ (so so) | 高品質な (good quality) | |
| ひどい (terrific) | 微妙 (subtle) | 適度に良い (moderately good) | すごい (amazing) | |

Figure 2 Candidate Expressions

2.2 Verbal Rating Scale Calibration

The aim of the scale calibration process was to select, among the previously identified 52 candidate items, an appropriate sub-set that best fit a number of criteria of interval-type measurement, as outlined below. We employed the method of equal appearing intervals (Thurstone 1954). To this purpose we asked a sample of 104 Japanese customers to sort the 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.*” Such procedure may be confusing to respondents since they mostly perform evaluations of objects such as brands or

products (=stimulus scaling), but not verbal expressions as in our study (=response scaling). To avoid misunderstanding and facilitate their task, respondents were thoroughly briefed by the researchers and given instructions together with the scenario shown in Figure-3.

Suppose you are discussing with friends about the quality of different brands and models of mobile phones. To express your opinion you may state that the quality of your favorite model or brand is "really good" or "superior," but you may call the quality of another less favorable model or brand being just "acceptable," and yet another one as "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 must NOT evaluate a specific model or brand, but evaluate each expression's meaning.

| | <i>This is the worst thing I could say about quality</i> | | | | | <i>This is the best thing I could say about quality</i> | | | | | |
|-------------------|--|--------------------------|--------------------------|--------------------------|--------------------------|---|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| | -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 3 Response Scaling

Following Thurstone, items should be selected so that they evenly span the entire attitudinal continuum. Item means or medians provide numerical representations of the item position, and standard deviation of item means indicate the amount of item-ambiguity. The smaller the standard deviation, the more consistently an item is rated within the calibration sample. To best respect the requirements of an interval-level scale we selected items for which mean distances were as equal as possible, and we discarded items with large standard deviations. This resulted into two alternative scales with respectively five and seven anchor items (Figure-4). Both scales are composed by items with fairly equal mean and/or median distances, and most standard deviations are inferior to 1 (Table-1). Both scales may be used to measure Japanese consumers' perceptions of product quality. Respondents can be simply asked to evaluate product quality by selecting one scale item that best reflects their opinion. So-gathered data can be used by researchers to investigate mean differences between groups of respondents, or to perform any other statistical operation requiring interval-type data.

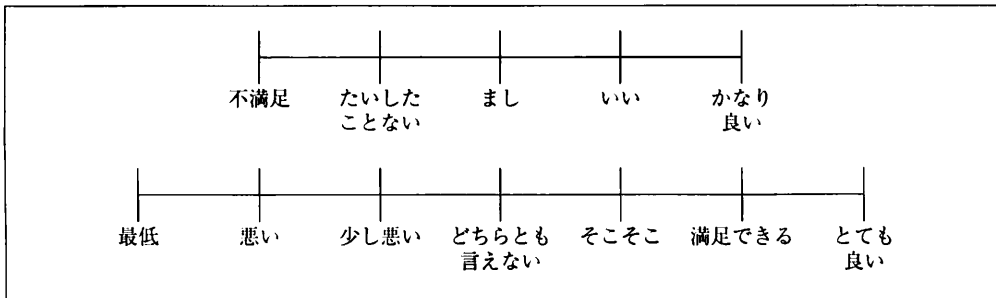


Figure 4 Five-Item and Seven-Item Verbal Rating Scales

Table 1 Psychometric Properties of the Five-Anchor and the Seven Anchor Scale

| | | Scale Anchor | | | | | | |
|------------------|-----------|--------------|------|------|------|-------|------|-------|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Five Item Scale | Std. Dev. | 2.56 | 4.55 | 6.45 | 8.54 | 10.48 | — | — |
| | Mean | 0.96 | 1.26 | 1.10 | 0.94 | 0.64 | — | — |
| Seven Item Scale | Std. Dev. | 1.29 | 2.77 | 4.34 | 5.88 | 7.28 | 8.68 | 10.24 |
| | Mean | 0.57 | 0.83 | 0.58 | 0.53 | 1.04 | 1.08 | 0.74 |

3. Conclusions

The main objective of this research was to develop a Japanese version of a new verbal rating scale to measure consumers' perceptions of product quality. Generally, verbal rating scales provide several advantages such as ease-of-explanation and respondents' familiarity with scale grades as advantages of verbal rating scales (Moxey and Sanford 1991; Rohrmann 2003). Furthermore, verbal rating scales may result in less skewed data, thus improving measurement precision (Westbrook 1980). Our analysis resulted into two different scales, a five-anchor and a seven-anchor scale. The five anchor scale may be more appropriate when researchers wish to evaluate a large number of objects such as product alternatives or different brands. Operating with fewer scale anchors may facilitate respondents' task. However, the seven anchor scale offers more answer alternatives which may make it the better choice when researchers want to differentiate between individuals. Future research should clear these points up. Second, the here developed verbal rating scale require formal validation. Future research should investigate the scales reliability or generalizability across different samples of respondents and furthermore investigate

its nomological validity. Finally, future research should also investigate predictive validity of the scale and show convergence and discriminant validity with other quality measurement. We are currently doing research to advance these fields.

References

- Buzzell, R. D. and B. T. Gale (1987), *The PIMS Principles: Linking Strategy to Performance*, The Free Press.
- Ishikawa, K. (1982), *Guide to Quality Control*, Tokyo, Asian Productivity Organization.
- Moxey, L. M. and A. J. Sanford (1991), "Context effects and the communicative functions of quantifiers: Implications for their use in attitude research", *Context effects in social and psychological research*, N. Schwarz and S. Sudman, New York, Springer.
- Rohrmann, B. (2003), "Verbal qualifiers for rating scales: Sociolinguistic considerations and psychometric data", available at www.psych.unimelb.edu.au/staff/rohrmann.html. Melbourne, University of Melbourne.
- Thurstone, L. L. (1927), "Three psychophysical laws", *Psychological Review* 34(6): 424-432.
- Thurstone, L. L. (1954), "The measurement of values", *Psychological Review* 61(1): 47-58.
- Westbrook, R. A. (1980), "A Rating Scale for Measuring Product/Service Satisfaction", *Journal of Marketing* 44(4): 68-72.