



UNIVERSITY  
OF WOLLONGONG  
AUSTRALIA

University of Wollongong  
**Research Online**

---

Faculty of Business - Papers

Faculty of Business

---

2017

# Optimal standard measures for marketing

John R. Rossiter

*University of Wollongong*, [jrossite@uow.edu.au](mailto:jrossite@uow.edu.au)

---

## Publication Details

Rossiter, J. R. (2017). Optimal standard measures for marketing. *Journal of Marketing Management*, 33 (5-6), 313-326.

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library:  
[research-pubs@uow.edu.au](mailto:research-pubs@uow.edu.au)

---

# Optimal standard measures for marketing

## **Abstract**

In this article, the author argues that marketing will not become a science until we agree on an optimal standard measure (OSM) for each of our major constructs. The case for OSMs is made by critically examining the leading alternative measures of four constructs used widely in marketing management ; corporate business reputation, corporate ethical reputation, customer satisfaction, and customer recommendation ; and showing how we might progress towards designing an OSM for each.

## **Disciplines**

Business

## **Publication Details**

Rossiter, J. R. (2017). Optimal standard measures for marketing. *Journal of Marketing Management*, 33 (5-6), 313-326.

# Optimal Standard Measures for Marketing

**John R. Rossiter**

Honorary Professorial Fellow

School of Management, Operations and Marketing

Faculty of Business

University of Wollongong

Wollongong NSW 2522

Australia

Telephone: + 62 4221 5660

E-mail: [john\\_rossiter@uow.edu.au](mailto:john_rossiter@uow.edu.au)

**Acknowledgement:** The author thanks Grahame Dowling, Faculty of Business, University of Technology Sydney, for his helpful comments on an earlier version of this paper.

## **Author Biography**

John R. Rossiter is Honorary Professorial Fellow in the School of Management, Operations & Marketing, Faculty of Business, University of Wollongong, Australia, and Visiting Professor of Marketing in the Schumpeter School of Business and Economics, Bergische University Wuppertal, Germany. His interests and publications are in marketing theory, buyer behaviour, advertising, and more recently social science measurement with particular applications to marketing.

# Optimal Standard Measures for Marketing

## ABSTRACT

The present author argues that marketing will not become a science until we agree on an optimal standard measure (OSM) for each of our major constructs. The case for optimal standard measures is made in this article by critically examining the leading alternative measures of four constructs used widely in marketing management – corporate business reputation, corporate ethical reputation, customer satisfaction, and customer recommendation – and showing how we might progress toward designing an OSM for each.

**Keywords:** Failure of psychometric measures, Multiple-item versus single-item measures, Optimal standard measures

## Summary Statement of Contribution

The author contributes to improving the validity of marketing knowledge by explaining why we cannot continue to use alternative measures of our major constructs and instead must choose an optimal standard measure of each of them.

## Optimal Standard Measures for Marketing

### ABSTRACT

The present author argues that marketing will not become a science until we agree on an optimal standard measure (OSM) for each of our major constructs. The case for optimal standard measures is made in this article by critically examining the leading alternative measures of four constructs used widely in marketing management – corporate business reputation, corporate ethical reputation, customer satisfaction, and customer recommendation – and showing how we might progress toward designing an OSM for each.

**Keywords:** Failure of psychometric measures, Multiple-item versus single-item measures, Optimal standard measures

## Optimal Standard Measures for Marketing

A question long debated is whether or not marketing is a science. We have journals called *Marketing Science* and the *Journal of The Academy of Marketing Science* but are they scientific in name only? A fundamental aspect in which marketing has *not* been scientific is *measurement*, and the main problem that prevents marketing from becoming a science is our readiness to accept alternative measures of the same construct. In the social sciences, including marketing, we accept new measures of a construct almost cavalierly, just as long as the new measure has ‘good psychometric properties’; the new measure then joins the set of alternative acceptable measures and researchers are free to use any of the measures in any given study and even to change the measure if they do a series of studies. Alternative measures are not accepted in the hard sciences, where the practice has always been to decide as quickly as possible on one measure as the standard.

Our acceptance of alternative measures is due to our wholesale adoption, ever since the appearance of Churchill’s (1979) famous *Journal of Marketing Research* article, of the *psychometric approach* to measure design. The lone critic of psychometrics has been the present author (see Author, 2002, 2005, 2008, 2011, 2013) and one of the major psychometric techniques he attacked was the procedure called *convergent validation*. Psychometrically trained researchers assume that if scores from two measures are significantly positively correlated – convergent – then the findings obtained from them will be the same, at least within the bounds of statistical significance. The present author, however, has recently found a mathematical disproof of this assumption in the

work of Carlson and Herdman (2012). The disproof will not be detailed here because it has been published as Author (2015) and is forthcoming as part of Author (2016) but it rests on the surprising fact that measures whose scores correlate even as highly as  $r = .90$ , which is about as convergent as you can get without using completely redundant measures, will produce widely divergent findings. The demonstration that different measures of the same construct do not produce equivalent findings – the failure of convergent validation – can mean only one thing: inevitably, that we must find the best measure of a given construct and use only that measure. Author (2016) calls this best measure the *optimal standard measure*. An optimal standard measure, or OSM, is the measure that a majority of experts in the field believe to be the most accurate measure of the construct, given that the construct *definition* has been agreed on beforehand.

But, to quote from *Hamlet*, ‘here’s the rub.’ All the so-called measurement experts in the social sciences are psychometricians. (As far as the present author is aware from published work, there are only two other anti-psychometricians in the social sciences apart from himself. The two are the late A.S.C. Ehrenberg, who was a hard-line anti-statistician, always used only single-item measures, and presumably held no truck with psychometrics, and Lars Bergkvist, lead author with the present author on the highly cited 2007 *JMR* article on when to use single-item measures and the only other researcher to publish anti-psychometrics articles.) Psychometricians seem uninterested in the hard intellectual work of defining constructs and then matching the content of the measure to the content of the construct definition, which is all that is needed to

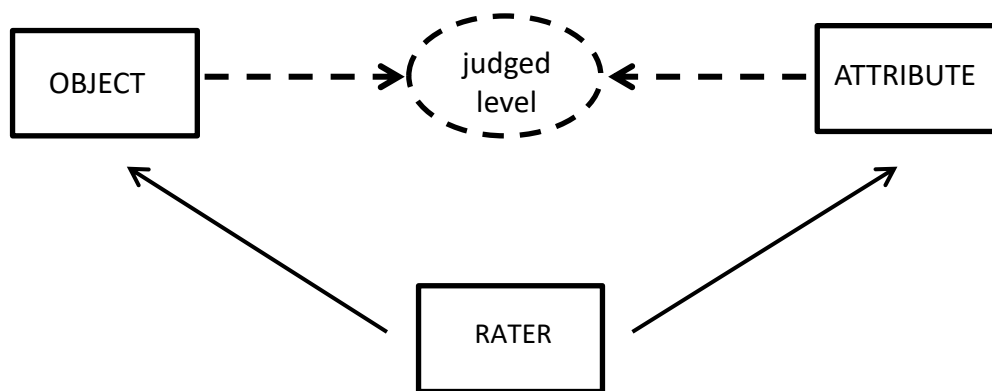
guarantee a valid measure. What psychometrically trained researchers typically do, instead, is write, or more often borrow, an approximate pool of items, administer them to a convenient pretest sample, and then run the scores through psychometric statistical programs under the foolish belief that the computer will tell them which items they should retain. Given that those who control our journals see no problem with this procedure, it is easy to see why researchers will not be keen to adopt a non-psychometric approach to measure design.

Nevertheless, as the biblical David might have said before setting out to slay Goliath, ‘somebody has to do it.’ Ehrenberg is most unfortunately no longer with us and Bergkvist has too much on his plate with teaching duties and the conducting of original research. So, as a senior marketing academic with his teaching and market research careers well behind him, it looks very much as though that ‘somebody’ is yours truly. The present author – who, having to mostly work alone, apologizes in advance for the unusually high number of self-citations – has already begun this task. During the past decade he has proposed an OSM for the construct of ‘market orientation’ (Author, 2012a), the construct of ‘e-service quality’ (Author, 2007, 2009), the construct of ‘brand love’ (Author, 2012b), and for the several constructs forming the well-known Technology Acceptance Model (Author & Braithwaite, 2013). In the present article he continues this task by proposing an OSM for four other important constructs in marketing management: corporate business reputation, corporate ethical reputation, customer satisfaction, and customer recommendation.



### *Construct definition primary*

The vital task prior to proposing an optimal standard measure is to propose a standard *definition* of the construct that we want to measure. All four of the constructs discussed in this article can be classified as *judgement constructs* in that they involve a judge, or *rater*, making a judgement about an *object* – a company, brand, or person as the case may be – in terms of a particular *attribute* or set of attributes. This construct type conforms to the present author’s object-attribute-rater model (Author, 2002), an updated depiction of which is shown in Figure 1. As can be seen, there is an object to be judged, an attribute to judge it on, and a rater *entity* to do the judging, and the rater’s task is to judge the *level* of the attribute as possessed by the object.



**Figure 1.** Object-Attribute-Rater model.

The article proceeds, for each of the four constructs, by first reviewing leading alternative construct definitions and their associated measures, and then proposing new construct definitions followed by a suggested optimal standard measure of each. Most of the measures reviewed were sourced from the two

widely used marketing measures *handbooks* (Bearden, Netemeyer, & Haws, Third Edition, 2011; and Bruner, Hensel, & James, 2005) and were published in major marketing journals such as the *Journal of Marketing (JM)*, the *Journal of Marketing Research (JMR)*, and the *Journal of Advertising (JA)* and are critiqued here for the very first time. Three popular practitioner measures are also critiqued.

Again, by way of an outline, the constructs examined in this article are corporate business reputation, corporate ethical reputation, customer satisfaction, and customer recommendation. Future directions are speculated about at the end of the article.

### **Corporate business reputation**

The construct of corporate business reputation is important to senior management, of course, and it is also important to marketing managers because they are usually entrusted with publicizing and protecting this reputation. Its importance is signalled by the emergence of a journal, *Corporate Reputation Review*, dedicated to this construct.

#### *Previous approaches*

The Bearden *et al.* handbook does not include measures of corporate reputation. The Bruner *et al.* handbook, however, offers two measures of corporate reputation, both very different. Also, there is one highly respected practitioner measure to be reviewed.

The first academic measure was designed by Sankar and Bhattacharya, hereafter S&B, and was published in *JMR* in 2001 (see Bruner, p. 54). Like the present approach discussed shortly, S&B separated corporate reputation into business reputation and social responsibility reputation, with the latter concerning matters of business ethics. The corporate reputation construct was called simply ‘company evaluation,’ and was defined in terms of the *evaluative beliefs* that the (unspecified) person holds about the company with regard to its business abilities. The problem with this definition is that it refers to the *antecedents* of corporate business reputation rather than to the reputation itself. The measure follows suit by assuming, with its five items, that there are five attributes that cause and form business reputation: manufacturing ability, technological innovativeness, product quality, customer service, and a wide range of products. With four out of the five attributes referring to products, it should be obvious that this measure would not be applicable to service companies, which now dominate over manufacturing companies in most Western countries. More subtly considered, the S&B measure assumes that everyone considers *only* these five attributes in arriving at a judgement of corporate business reputation *and* that everyone undertakes this multi-attribute mental computation rather than making a straightforward global judgement of reputation.

The second academic approach to defining and measuring corporate business reputation was that of Goldsmith, Lafferty, and Newell and was published in the *Journal of Advertising* in 2001 (see Bruner, p. 53). Goldsmith and colleagues, hereafter GL&N, defined the construct as a person’s *general*

*opinion* of the company, which does imply a global judgment of reputation, but, like S&B, they used multiple items to measure it. The items were ‘favourable-unfavourable,’ which would have sufficed on its own as a single-item measure; ‘good-bad,’ which refers more to a moral or ethical judgement; and the construct-irrelevant item ‘satisfactory-unsatisfactory,’ and so the addition of these last two items *de-validates* the measure by taking it ‘off construct.’ Also, as with the previous measure, the rater entity in GL&N’s definition is referred to vaguely as ‘a person,’ instead of specifying that the rater be a concerned stakeholder.

The most favoured industry conceptualization of corporate business reputation is the ranking of ‘most admired’ companies, initiated by *Fortune* magazine in the early 1980s for U.S. companies and soon imitated by business publications in Asia and Britain for companies based in their region. In each case, only the ranking is reported – as though it were a single-item measure. Little-known, however, is that the corporate reputation ranking systems, not unlike worldwide university ranking systems, are multiple-item measures based on the sum of ratings on different attributes (for details of these differing attributes, see Fombrun, 1998) which means that you cannot legitimately compare the rankings of a multiregional or global company across regional areas. (Only the *The Financial Times* survey, covering companies based in the U.K. and Europe, takes a direct ranking of what the publication calls the ‘most respected’ companies.) The problem with derived ranking measures is that, like S&B’s measure earlier, they assume, unreasonably, that raters – business managers in this case – consider only these particular attributes before making a business

reputation judgement and, moreover, that business managers actively compute this judgement by summing or averaging the ratings.

The measures also are not informative about how to *improve* a company's business reputation because the measures presume different antecedents or 'causes' of the reputation. (The *FT* survey is the notable exception here: it asks for the overall ranking of respected companies first, and then asks the respondents to name the main reason for the ranking without biasing the ranking itself.) The managerially informative aspect will be taken up in conjunction with the present approach.

### *The present approach*

Corporate business reputation is primarily relevant to two different rater entities – potential investors and prospective employees. The constructs differ for these two rater entities because the *attribute* differs with the rater entity difference. Potential investors are mainly interested in the company's medium- to long-run financial prospects, so what matters to them is how good they perceive the company to be as a current investment target. Prospective employees, on the other hand, defined as those capable of gaining a job in that company's industry, are interested mainly in whether it would be a good company to work for. This line of argument suggests that corporate business reputation should be separated into two constructs that we can call, respectively, the company's investment appeal and the company's job appeal.

The two constructs are what Author (2002) called 'doubly concrete,' consisting of a clear single object (the company) judged on a single unambiguous

attribute (investment appeal and job appeal, respectively). This means that, in turn, each construct must be measured *single item* if it is to function as a global belief in the rater's mind (see also Bergkvist & Author, 2007).

Suggested single-item measures of company investment appeal and company job appeal are provided in Table 1. The new measures employ what the present author calls the *level-free forced-choice binary* format, which every market researcher should consider. Two articles in the *International Journal of Market Research*, previously the house journal of the British Market Research Society, describe this new belief measure. The article by Dolnicar, Author, and Grün (2012) demonstrates that the forced-choice binary measure is much more valid and stably reliable than either the 'pick any' measure of brand-attribute beliefs that Ehrenberg used in his buyer behaviour research or the 'multi-point,' typically 7-point, numerical belief measure that academics use, and the article by Author, Dolnicar and Grün (2015) explains how the new measure works and why it works so well – an explanation, to get a bit technical here, that involves the measure's automatic capture of individual raters' heterogeneous thresholds for answering affirmatively.

**Table 1.** Corporate business reputation separated into two constructs with a suggested single-item measure of each.

---

**Investment Appeal:** The potential investor's (rater entity) belief that Company X's stock (object) is currently a good investment (attribute).

**Measure:** Do you believe that Company X, at present, is a good investment? Yes No

Job Appeal: The potential industry employee's (rater entity) belief that Company X (object) would be a good company to work for (attribute).

*Measure*: Would you say that Company X is a company that you would like to have your job with? Yes No

---

If a ranking of companies is wanted – and ranking only makes sense within industry – then a simple tally of the ‘yesses,’ taken with a reasonably large and random sample of target raters, would yield a valid and statistically reliable ordering.

Market researchers (and perhaps managers) are likely to argue that the single-item measures are inadequate because they do not reveal *why* the rater would or would not invest, or would or would not like to work for, their company. The best answer to this type of question is that multiple-item measures such as those reviewed above *presume* the reasons by including them as antecedents of overall reputation. It is far more valid – although academic researchers, especially, are reluctant to do this – to ask an *open-ended* follow-up question about *why* the rater answered yes or answered no. Categorization (coding) of the first couple of reasons given by each person would provide valid indications of the main positive and negative factors influencing the rater's judgement, and also indicate how the company's investment-appeal or job-appeal reputation might be improved.

Whereas the separation of corporate business reputation into two constructs can be debated along with the best wording of their respective measures, this is the exactly the type of debate that the present author is trying to encourage. Otherwise, we will never progress toward deciding on optimal standard measures.

### **Corporate ethical reputation**

Clearly growing in importance, in Western countries at least, is the company's *ethical reputation* as distinct from its business reputation. It is difficult to assess just how important corporate ethical reputation is in a purely economic sense but is not difficult to defend its importance from a modern marketing perspective. For a growing number of potential customers and for prospective employees as well, corporate ethical reputation might well be a 'knockout' consideration that overrides a good business reputation. Coal-mining companies and liquor and tobacco companies are examples, as are manufacturing companies reputed to use 'sweatshop' labour.

### *Previous approaches*

Bearden's handbook lists four measures related to business ethics (pp. 446-455) and Bruner's handbook lists two (pp. 55, 183). Reviewed here are two of the most different measures which might be wrongly regarded as equivalent. One of the measures is Hunt, Wood and Chonko's *corporate ethics scale*, published in *JMR* in 1989 (see Bearden, pp. 451-452). The problem with HW&C's conceptualization of corporate business ethics is that it consists of three



parts, two of which are completely ‘off construct.’ The first part is the employee’s belief about whether their managers are acting ethically (which does fit the construct) but then the researchers added – or more likely inferred *ex post facto* from their measure – two other parts: the employee’s belief about whether their managers are concerned about ethical issues, and the employee’s belief about whether he or she, not the manager, would be punished for unethical behaviour (both of which are irrelevant to the construct of corporate ethical reputation). However, the reader should notice that even the *first* part of Hunt et al.’s conceptualization is ‘off construct’ because it refers to whether the *employee*, an inappropriate rater entity, considers the company, or at least its managers, to be behaving unethically, when surely it is *outsiders’ beliefs* that matter most.

A different measure is offered by Sankar and Bhattacharya in their 2001 *JMR* article alongside their business reputation measure reviewed above (see Bruner, p. 183). In that article, S&B measured a construct that they called *corporate social responsibility – personal support*. Their measure differs from Hunt’s in that it covers *specific* ethical issues – 10 of them in fact, such as ‘equal-opportunity employment’ and, contradictorily enough, ‘affirmative action.’ However, S&B make the mistake of measuring whether the individual *respondent*, rather than the *company*, supports each of the issues and thus is very badly off-construct.

Neither of the two measures of corporate ethical reputation has sufficient construct-to-measure validity. Also, they are completely different and would necessarily produce completely different findings.

*The present approach*

The present approach to defining and measuring corporate ethical reputation departs entirely from all previous approaches. The conceptual argument here is that the company needs to be seen to have made only *one* recent ethical breach to be saddled with a reputation of being *unethical*. A good example of a damaging one-off event would be Volkswagen's admitted attempt to bypass its cars' fuel emission recording devices.

Also, it doesn't really matter what the public at large believes about the company. The most marketing-relevant rater entity for the construct of corporate ethical reputation would be *potential customers* – essentially those who could buy the company's product or patronize its service next time they are 'in the market.'

The corporate ethical reputation construct, given these arguments, can therefore be measured with a single item – but a different type of item than the overall judgment item used for corporate reputation. A suggested construct definition and measure are shown in Table 2.

**Table 2.** Corporate ethical reputation defined and a suggested single-item measure.

---

**Corporate Ethical Reputation:** The potential customer's (rater entity) belief that the company (object) engages in what the customer believes to be an unethical production or selling practice (attribute).

**Measure:** Company X is involved in an unethical production or selling practice. Yes No

---

It is possible, too, that an ethical breach could be taken as negative for potential *investors* because of the likelihood of the company having to pay costs or damages. If so, this is best left as an option for an open-ended response following the investor's corporate business reputation measure suggested earlier.

Whereas the company was the focus of – the object in – the corporate reputation constructs, the analysis now turns to constructs in which the *customer* is the focus. The two constructs examined are customer satisfaction and customer recommendation.

### **Customer satisfaction**

The construct of customer satisfaction gets to the very heart of 'the marketing concept': companies are supposed to satisfy customers as *the* means of ensuring long-run profit and guaranteeing survival. Customer satisfaction, however, is measured in various ways, differing mostly in terms of how 'satisfaction' is defined.

#### *Previous approaches*

Defining the attribute of satisfaction is the main conceptual problem. In the previous literature, satisfaction has been defined in three different ways: firstly, as a unipolar attribute ranging from zero satisfaction up to some extreme positive amount including, according to some theorists, 'delight'; secondly, as a bipolar attribute ranging from negative dissatisfaction to positive satisfaction; and, lastly, as the *smallness* of the 'gap' between what the customer is seeking

and what the company is perceived to be delivering. Notice that the first and second definitions see satisfaction as absolute, whereas the last definition sees it as relative – relative to the customer’s ideal. You should easily see that the three different definitions and their associated measures are going to produce different findings: a low score on the unipolar measure means that the customer is only slightly satisfied, on the bipolar measure it means that the customer is actually *dissatisfied* (a result often hidden by scoring the lower extreme as a 1 on a 7-point scale when it should be scored as –3), and on the gap measure a low score means the opposite in that with little or no gap between actual and ideal the customer should be *very* satisfied. A side point about the gap measure, were it to be accepted as the standard, is that it can be possibly unethically ‘gamed’ by companies as in the strategy known as ‘underpromise but overdeliver.’

The world’s leading practitioner measure is the American Customer Satisfaction Index, the ACSI, previously called the Customer Satisfaction Barometer when pioneered in Sweden and Norway. The ACSI is unusual for a practitioner measure in that it was designed by marketing academics – using the psychometric approach – and achieved notice when published in the *Journal of Marketing* (Fornell, Johnson, Anderson, *et al.*, 1996) which in the U.S. is the only marketing management journal that has any sizeable industry subscriber base. Although the ACSI comes up with a single absolute percentage satisfaction score, the ‘black box’ for this measure (exact items exposed in *Wikipedia* some years ago) reveals that it is computed as an impossible mixture of scores on three different types of satisfaction item: an absolute dissatisfaction-satisfaction item (bipolar), a relative satisfaction-satisfaction item based on falling short of vs.

exceeding expectations (bipolar), and a ‘gap’ item involving a comparison of delivered satisfaction to a self-defined ideal (which is unipolar). Only the first item was necessary, and combining its scores with those from the other two items is about as accurate as measuring a person’s height by using a tape measure then adding to it the person’s own estimate and your own guess as well.

Two further problems cloud the overall ACSI score. Although two of the items are bipolar, all three items are scored unipolar as 1 to 10, which obscures dissatisfaction. Also, the ACSI survey is conducted only with the company’s own current customers and misses potential and dissatisfied ex-customers.

The measure designers of the ACSI – as do all other measure designers who use multiple-item measures – committed another psychometric mistake (see Author, 2013). They defined customer satisfaction as a ‘latent’ construct that causes the responses to the items thus, technically speaking, defining and measuring it as ‘reflective’ when, if anything, it is ‘formed’ or, as argued next, is ‘doubly concrete.’

### *The present approach*

A good argument can be made that *customer satisfaction* is a construct consisting of a single concrete object (the branded product or service), a single concrete attribute (bipolar: positive satisfaction through negative dissatisfaction), and for which the rater entity should be three types of customer (potential new customers, current customers, and recently defected customers). There is no good reason why customer satisfaction-dissatisfaction cannot be measured single item and there is a very good reason for doing so. This reason is

that satisfaction-dissatisfaction obviously functions as a *single belief or attitude* arising in the potential customer's, the current customer's, or the ex-customer's mind. The three rater entities mean that you have to split customer satisfaction into *three* constructs – *new-customer expected* satisfaction-dissatisfaction, *current-customer* satisfaction-dissatisfaction, and *defected-customer* satisfaction-dissatisfaction. The measurement results of each construct have to be analyzed, and managed, separately.

The present author also argues that the *answer options* for customer satisfaction measures should consist of verbal options, rather than the usual numerical ones; this is so the answers will have clear implications for managerial action.

Suggested definitions and measures of the three customer satisfaction constructs are provided in Table 3. The present author's square-bracketed recommendations accompanying each definition should be verified by conducting industry-specific qualitative research up front and the measures modified if so indicated.

**Table 3.** Customer satisfaction separated into three constructs according to the rater entity, each measured single-item with verbal answer options.

---

Potential customers' satisfaction: The potential customer's (rater entity) *expectation* of how satisfied or dissatisfied they would be (attribute) with the company's product or service (object). [A hypothetical question, obviously, but one that is very important for potential customers to answer.]

Measure: Imagine that you were to become a customer of Company X... Based on what you know or might have heard about Company X, how satisfied or dissatisfied would you *expect* to be with that company's products or services? Very satisfied; Satisfied enough; Really cannot say; Somewhat dissatisfied; Very dissatisfied (no need to assign numerical scores because cross-tabulation is would be easier to understand, but if you want numbers the answers should be scored + 3, + 1, 0, -1, - 3, with extra weighting for extreme scores as per Likert, 1932).

Current customers' satisfaction: The current customer's (rater entity) rating of how satisfied or dissatisfied they were with the *last* purchase from or transaction with (attribute) the company (object). [The most recent purchase or transaction carries by far the most weight in determining current satisfaction.]

Measure: Think back to your most recent purchase or transaction with Company X... Would you say that you were: Very Satisfied; Satisfied enough and will continue with the company; Haven't really thought about it – neither satisfied nor dissatisfied; Dissatisfied enough to think about going to another company; So dissatisfied that you will definitely not buy from or use that company again (same comment as above about scoring).

Ex-customers' satisfaction: The defected customer's (rater entity) rating of how satisfied or dissatisfied they were with the most recent one or two purchases from or transactions with (attribute) the company (object). [Customers, according to the marketing textbooks, should give the company only one chance to under-deliver but, perhaps for reasons of cognitive dissonance from a personally committed choice, or from simple laziness or inertia, they often give the company two chances to do so before defecting.]

Measure: I hope you don't mind my calling but I am Bill Jones, representing Company X, and our records indicate that you may have decided to no longer be one of our customers. Is that correct? IF SOME OTHER REASON IS GIVEN SUCH AS MOVING OUT OF THE AREA, THEN RECORD THIS FACT, APOLOGIZE, THANK, AND TERMINATE. Was it dissatisfaction with one of our products or with something that we did that caused you to leave? IF YES: Could you please tell me more about that. PROBABLY USEFUL TO RECORD THE NATURE OF THE COMPLAINT, THOUGH THE ESSENTIAL ANSWER OPTIONS TO RECORD HERE ARE: Yes No (numerical scoring not required). IF NO: THANK AND TERMINATE.

---

The wording of the three measures should be carefully pretested with a small sample (of say 10 or so individuals) drawn from each rater-entity group. The rather painstaking yet colloquial wording of the three single-item measures is something that you would be taught in a market research firm but would be highly unlikely to be learned during a sheltered academic career.

An optimal standard measure of customer satisfaction – and there are three OSMs to be designed and agreed on here – would follow the structure and meaning, if not the exact wording, of the above.

### **Customer recommendation**

Customer recommendation is the other customer-focused construct highly regarded in marketing management at present. The history of this construct is briefly reviewed and then the focus shifts to the most popular practitioner measure.



### *Previous approaches*

The *customer recommendation* construct is also known in the consumer marketing industry as *customer advocacy* (see, e.g., Spooner, 2012). Previously in the academic literature, customer recommendation was called *word-of-mouth* (see Bruner, p. 647, for a naive and redundant 3-item measure of recommendation by word-of-mouth published in *JM* in 1999) and, even earlier than that, came from sociology as the construct called *interpersonal influence* (a thoughtful Everett Rogers-based measure of which was designed by King and Summers in 1970 and incorrectly modified in 1986 by a later academic researcher and published in the same journal, *JMR*, with both measures reproduced in Bearden, pp. 99, 100). But by far the most widely used measure of customer recommendation today is the practitioner-designed single-item measure published in the *Harvard Business Review* (Reichheld, 2003) called Net Promoter.

The Net Promoter measure delivers a score commonly known as the Net Promoter Score, or NPS. The single item measuring the NPS is: ‘How likely are you to recommend Brand X to a friend or colleague?’ Respondents – all of them current customers of Company X – answer this question on an 11-point scale ranging from 0 = ‘not at all likely’ through 10 = ‘extremely likely.’ In scoring the answers, respondents who give a likelihood rating of 9 or 10 are designated as Promoters; those who rate their likelihood as 0 through 6 are designated as Detractors; while those giving a 7 or 8 – mystifyingly enough, since it is these moderately positive customers who are likely to be the most important marketing target – are deleted from the NPS computation. The Net Promoter

Score is calculated as %Promoters minus %Detractors. For example, if the survey shows 50% Promoters, 40% Detractors, and 10% deletes, then the brand's NPS = 50 – 40 = 10. NPS is thus a net percentage difference score; for example, the 10, or +10%, signifies a predominance of 'promoters' and scores as low as –15 are reported for some brands which have a predominance of 'detractors.' Two of the highest NPS scores ever recorded – again among the company's own customers, remember – were 81% net promoters for Harley-Davidson motorcycles and 73% for Amazon.com (see Creamer, 2006). Most companies do not achieve anywhere near these numbers; Avis, for example, which presumably 'tries harder,' in 2006 achieved a net of only 28% promoters among its customers.

Despite its ready adoption by industry, Net Promoter has a number of obvious content-validity problems. First, Net Promoter measures *intention* to recommend, not actual recommendations, and there is often a big gap between intention and behavior, especially for 'high involvement' behaviors – those for which others' recommendations and disrecommendations are presumably most important. Second, individuals are likely to give *negative* as well as positive recommendations about a given brand (East, Hammond, and Wright 2007) and Net Promoter does not measure disrecommendation, that is, *negative* word-of-mouth. (The label 'detractors' is misleading because 0-to-6 scorers on Net Promoter are merely non-likely through moderately likely intenders to *positively* recommend the brand, so it is entirely unreasonable to presume that they will speak out *against* the brand.) Fourth, as Mangold, Miller, and Brockway (1999) have noted, opinion-givers quite often recommend a brand that would suit the *other person*, the opinion seeker, rather than themselves.

It should be mentioned that the Net Promoter rating scale has been used (at least in Australia) to measure the different and previously considered construct of *customer satisfaction*. This measure uses a different question along the lines of: ‘On a scale of 1 to 10, how would you rate my answer to your inquiry?’ The present writer’s younger son recently worked for an ‘inbound’ telephone-based customer service company that paid its customer service employees a bonus if they achieved promoter-level (9 or 10) ratings when the customer is asked for a rating at the end of the call. Calls are monitored on a random basis, preventing the temptation to nudge a little if the customer answers ‘8’!

### *Recommended approach*

The problems with customer recommendation measurement were largely solved by the U.K.’s Robert East and colleagues (2007, 2008; Uncles, East, and Lomax 2010). East and colleagues measured, for the particular product or service category, individuals’ recalled frequency in the past 6 months, of giving *and* receiving of recommendations and disrecommendations (thus four questions instead of Net Promoter’s single question). They then asked about the brand for which the person last gave positive advice, the brand for which the person last gave negative advice, the brand for which the person last received positive advice plus the strength of the recommendation, and the brand for which the person last received negative advice plus the strength of the counter-recommendation (six further questions). Overall, East et al. designed a 10-item measure in which all the items were independently necessary. A useful extension requiring four

more questions, thus 16 items in total, was suggested by market research practitioner Lyndall Spooner (2012). The extension is to ask the opinion-giver whether the receiver *acted* on the recommendation or on the disrecommendation, and likewise ask the opinion-recipient whether he or she acted on the positive or negative advice.

Notice that the 16-item measure is not a multiple-item measure in the conventional sense; rather, it is simply a series of *single-item* measures. If customer recommendations are as important as marketers claim them to be, then there is a good case for nominating this lengthy measure as the optimal standard measure.

## Conclusions

If marketing is to progress as a science, market theorists and researchers must agree on an *optimal standard measure* for each of our main constructs, just as is done in the hard sciences. We need *optimal* measures because otherwise we can trust neither our findings nor the statistical significance tests based on them. We need *standard* measures because, if the measure is changed, we are unable to legitimately compare findings and unable to properly replicate findings. The logical conclusion is *optimal standard measures*. Unfortunately, researchers in the social sciences remain ignorant of this argument and continue to accept alternative measures of the same construct, provided that the measures produce positively correlated scores. This may be termed the ‘near enough is good enough’ approach to measurement. However, in any science, near enough

is *not* good enough, and ‘near enough’ measures must not continue to be accepted.

Whereas most researchers to whom the present author has presented these arguments do not disagree with the need for optimal standard measures, they frequently object that the task is too difficult and that it is all too ‘quixotic’ to try. Those discouraged should consider the lifetime work of the buyer behaviourist and remarkably anti-statistics statistician, Andrew Ehrenberg (see Ehrenberg, 1972, for the beginnings of the buyer behaviour program that he followed consistently throughout his lifetime and see Ehrenberg, 1975, 1982, for his practitioner-friendly and stunningly sensible views on statistical analysis and data presentation). Ehrenberg’s research program is unique in marketing for using consistent definitions and measures throughout (and for the complete absence of statistics other than simple mean averages and mean deviations). Ehrenberg’s main constructs – unit market share, market penetration, repeat buying rate, and brand-attribute beliefs – are all of the basic or ‘doubly concrete’ type that require only a single-item measure.

The biggest danger in industry appears to be multiple-item measures masquerading as single-item measures. Two of the measures reviewed in this article, *Fortune’s* ‘most admired’ companies and the American Customer Satisfaction Index, are examples. Industry managers need to get their market research personnel to look into the ‘black box’ behind these measures to decide whether the multiple items are really necessary or whether a well thought out single-item measure would do a more accurate job. If the measure is to be used internally, then only an optimal measure is needed. But if companies want to

'benchmark' themselves against other companies, then the industry's market researchers need to get together as an industry body and design an optimal measure that can also serve as the standard measure.

Again, the purpose of this article is to urge researchers in marketing to begin the task of defining our major constructs carefully and agreeing on an optimal standard measure of each. The present author does not claim to have all the answers and is merely pointing out what we must do to raise marketing to the status of a legitimate science.

**REFERENCES**

- Author (2002). The C-OAR-SE procedure for scale development in marketing. *International Journal of Research in Marketing*, 19(4), 305-335.
- Author (2005). Reminder: a horse is a horse. *International Journal of Research in Marketing*, 22(1), 23-25.
- Author (2007). Toward a valid measure of e-retailing service quality. *Journal of Theoretical and Applied Electronic Commerce Research*, 2(3), 36-48.
- Author (2008). Content validity of measures of abstract constructs in management and organizational research. *British Journal of Management*, 19(4), 380-388.
- Author (2009). ER-SERVCOMPSQUAL: a measure of e-retailing service components quality. *Service Science*, 1(4), 212-224.
- Author (2011). Marketing measurement revolution: the C-OAR-SE method and why it must replace psychometrics. *European Journal of Marketing*, 45(11/12), 1589-1600.
- Author (2012a). Further comment on “market orientation.” *Australasian Marketing Journal*, 20(1), 108-112.
- Author (2012b). A new C-OAR-SE-based content-valid and predictively valid measure that distinguishes brand love from brand liking. *Marketing Letters*, 23(3), 905-916.
- Author (2013). Scientific progress in measurement theory? *Academy of Marketing Science Review*, 3(3), 171-179.
- Author (2015). Commentary: why you must use my C-OAR-SE method. *Australasian Marketing Journal*, 23(3), 259-260.

- Author (2016). How to use C-OAR-SE to design optimal standard measures. *European Journal of Marketing*, 50, in press.
- Author, & Braithwaite, B. (2013). C-OAR-SE-based single-item measures for the two-stage Technology Acceptance Model. *Australasian Marketing Journal*, 21(1), 30-35.
- Author, Dolnicar, S., & Grün, B. (2015). Why the level-free forced-choice binary measure of brand benefit beliefs works so well. *International Journal of Market Research*, 57(2), 239-256.
- Bearden, W.O., Netemeyer, R.G., & Haws, K.L. (2011). *Handbook of Marketing Scales, Third Edition*. Thousand Oaks, CA: Sage.
- Bergkvist, L., and Author (2007). The predictive validity of multiple-item vs. single-item measures of the same constructs. *Journal of Marketing Research*, 44(2), 175-184.
- Bruner, G., Hensel, P.J., & James, K.E. (2005). *Marketing Scales Handbook, Volume 4: A Compilation of Multi-Item Measures for Consumer Behavior and Advertising*. Mason, OH: Thomson/South-Western and the American Marketing Association.
- Carlson, K.D., and Herdman, A.O. (2012). Understanding the impact of convergent validity on research results. *Organizational Research Methods*, 15(1), 17-32.
- Churchill, G.A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, 16(1), 64-73.



- Creamer, M. (2006). Do you know your score? Your business rides on how consumers answer there is one crucial question. *Advertising Age*, July 3, pp. 1, 24.
- Dolnicar, S., Author, & Grün, B. (2012). 'Pick any' measures contaminate brand image studies. *International Journal of Market Research*, 54(6), 821-834.
- East, R., Hammond, K., and Lomax, W. (2008). Measuring the impact of positive and negative word of mouth on brand purchase probability. *International Journal of Research in Marketing*, 25(3), 215-224.
- East, R., Hammond, K., and Wright, M. (2007). The relative incidence of positive and negative word of mouth: a multi-category study. *International Journal of Research in Marketing*, 24(2), 175-184.
- Ehrenberg, A.S.C. (1972). *Repeat Buying: Theory and Applications*. Amsterdam: North-Holland/Elsevier.
- Ehrenberg, A.S.C. (1975). *Data Reduction: Analysing and Interpreting Statistical Data*. London: Wiley.
- Ehrenberg, A.S.C. (1982). *A Primer in Data Reduction: an Introductory Statistics Textbook*. Chichester: Wiley.
- Fombrun, C.J. (1998). Indices of corporate reputation: an analysis of media rankings and social monitors' ratings. *Corporate Reputation Review*, 1(4), 327-340.
- Fornell, C., Johnson, M.D., Anderson, E.W., Cha, J., and Bryant, B.E. (1996). The American Customer Satisfaction Index: nature, purpose, and findings. *Journal of Marketing*, 60(4), 7-18.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22 (whole of no. 140), 5-55.

Mangold, W.G., Miller, F., and Brockway, G.R. (1999). Word-of-mouth communication in the service marketplace. *Journal of Services Research*, 13(1), 73-89.

Reichheld, F.F. (2003). The one number you need to grow. *Harvard Business Review*, 81(December), 46-54.

Spooner, L. (2012). 'The cost of customer advocacy.' *Research News*, September, pp. 31-32.

Uncles, M., East, R., and Lomax, W. (2010). Market share is correlated with word-of-mouth volume. *Australasian Marketing Journal*, 18(3), 145-150.

*Wikipedia*. ACSI score calculation. Accessed July 24, 2012.