University of Wollongong

Research Online

Faculty of Business - Papers (Archive)

Faculty of Business and Law

1-1-2016

How to use C-OAR-SE to design optimal standard measures

John R. Rossiter

University of Wollongong, jrossite@uow.edu.au

Follow this and additional works at: <https://ro.uow.edu.au/buspapers>

Part of the [Business Commons](#)

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

How to use C-OAR-SE to design optimal standard measures

Abstract

Purpose: This paper aims to extend Rossiter's C-OAR-SE method of measure design (IJRM, 2002, p. 19, p. 4, pp. 305-335; EJM, 2011, p. 45, p. 11, p. 12, pp. 1561-1588) by proposing five distinct construct models for designing optimally content-valid multiple-item and single-item measures.

Design/methodology/approach: The paper begins by dismissing convergent validation, the core procedure in Nunnally's (1978) and Churchill's (1979) psychometric method of measure design which allows alternative measures of the same construct. The method of dismissal is the mathematical demonstration that an alternative measure, no matter how highly its scores converge with those from the original measure, will inevitably produce different findings. The only solution to this knowledge-threatening problem is to agree on an optimal measure of each of our major constructs and to use only that measure in all future research, as is standard practice in the physical sciences. The paper concludes by proposing an extension of Rossiter's C-OAR-SE method to design optimal standard measures of judgment constructs, the most prevalent type of construct in marketing.

Findings: The findings are, first, the mathematical dismissal of the accepted practice of convergent validation of alternative measures of the same construct, which paves the way for, second, the proposal of five new C-OAR-SE-based construct models for designing optimal standard measures of judgment constructs, three of which require a multiple-item measure and two of which a single-item measure.

Practical implications: The common practice of accepting alternative measures of the same construct causes major problems for the social sciences: when different measures are used, it becomes impossible, except by remote chance, to replicate findings; meta-analyses become meaningless because the findings are averaged over different measures; and empirical generalizations cannot be trusted when measures are changed. These problems mean that we cannot continue to accept alternative measures of the constructs and that, for each construct, an optimal standard measure must be found.

Originality/value: The ideas in this paper, which have untold value for the future of marketing as a legitimate science, are unique to Rossiter's C-OAR-SE method of measure design.

Disciplines

Business

Publication Details

Rossiter, J. R. (2016). How to use C-OAR-SE to design optimal standard measures. *European Journal of Marketing*, 50 (11), 1924-1941.

How to use C-OAR-SE to design optimal standard measures

John R. Rossiter

Faculty of Business, University of Wollongong, Australia

Schumpeter School of Business and Economics, Bergische University Wuppertal, Germany

john_rossiter@uow.edu.au

December 11, 2015

How to use C-OAR-SE to design optimal standard measures

Abstract

Purpose – This article extends Rossiter’s C-OAR-SE method of measure design (*IJRM*, 2002, 19, 4, 305-335; *EJM*, 2011, 45, 11/12, 1561-1588) by proposing five distinct construct models for designing optimally content-valid multiple-item and single-item measures.

Design/methodology approach – The article begins by dismissing convergent validation, the core procedure in Nunnally’s (1978) and Churchill’s (1979) psychometric method of measure design which allows alternative measures of the same construct. The method of dismissal is the mathematical demonstration that an alternative measure, no matter how highly its scores converge with those from the original measure, will inevitably produce different findings. The only solution to this knowledge-threatening problem is to agree on an optimal measure of each of our major constructs and to use only that measure in all future research, as is standard practice in the physical sciences. The article concludes by proposing an extension of Rossiter’s C-OAR-SE method to design optimal standard measures of judgment constructs, the most prevalent type of construct in marketing.

Findings – The findings are, firstly, the mathematical dismissal of the accepted practice of convergent validation of alternative measures of the same construct, which paves the way for, secondly, the proposal of five new C-OAR-SE-based construct models for designing optimal standard measures of judgment constructs, three of which require a multiple-item measure and two of which a single-item measure.

Practical implications – The common practice of accepting alternative measures of the same construct causes major problems for the social sciences: when different measures are used it becomes impossible, except by remote chance, to replicate findings; meta-analyses become

meaningless because the findings are averaged over different measures; and empirical generalizations cannot be trusted when measures are changed. These problems mean that we cannot continue to accept alternative measures of our constructs and that, for each construct, an optimal standard measure must be found.

Originality/value – The ideas in this article, which have untold value for the future of marketing as a legitimate science, are unique to Rossiter's C-OAR-SE method of measure design.

Keywords Psychometrics; Convergent validation; Construct-to-measure validity; Object-on-attribute judgment constructs; C-OAR-SE construct models for judgment constructs; Optimal standard measures

1. Introduction

In the opening sentence in the abstract of my *EJM* article on the C-OAR-SE method (Rossiter, 2011a, p. 1561) I drew attention to the fact that: “New measures in marketing are invariably created by using a psychometric approach based on Churchill’s ‘scale development’ procedure.” The new measure might be an *alternative* measure of the same construct, such as Kohli and Jaworski’s measure of MARKET ORIENTATION, which is supposedly interchangeable with Narver and Slater’s earlier measure (see Rossiter, 2012); or an *adapted* measure of the same construct, such as the many different items employed in adaptations of McKenzie and Lutz’s original measure of ATTITUDE TOWARD THE AD (see Bergkvist and Rossiter, 2009); or a *shortened* measure of the same construct, such as Zaichkowsky’s 10-item version of her original 20-item measure of PERSONAL INVOLVEMENT (see Rossiter, 2002); or the substitute measure might be a so-called *proxy* measure of the construct, as is common practice in marketing science.

Researchers brought up on the conventions of psychometrics claim the new measure to be a valid – and therefore an acceptable alternative – measure of the construct because scores from it are shown empirically to be correlated significantly with scores from the original measure. This procedure is known as “convergent validation,” Cronbach and Meehl’s (1955) concept that was famously operationalized by Campbell and Fiske (1959) as “multitrait-multimethod” analysis and given major impetus in psychology by Nunnally’s textbook on psychometrics, especially the 1978 edition. This second edition coincided with social science researchers’ wholesale switch during the late 1970s and early 1980s to using “canned” statistical software, mainly SPSS, for all data analysis. This handy set of statistical programs, usable without consulting the explanatory manual, made the psychometric techniques of correlational analysis, factor analysis, and the computation of coefficient alpha available to all as an easy way to make their often poorly designed measures look sound. In

marketing, Churchill's (1979) *JMR* article brought psychometrics to the masses, and in that article convergent validation and MTMM analysis are described (on pp. 70-72) as the *only* way to validate a new measure.

The logical flaws in convergent validation were exposed in my 2011 *EJM* article (on pp. 1562-1565). The overall problem is that it is logically illegitimate to argue that a measure validly represents the construct on the basis of its relationship to *other* measures (to put the problem simply, it is not logical to argue "from the outside to the inside"). As well, there are two further problems with convergent validation. One problem is that you cannot convergently validate the *first* measure of a construct, obviously, if there is no prior measure to relate it to. The other problem is that when there are two convergently valid measures of the same construct, you cannot tell from the convergent correlation which of them is the better measure.

The only answer to the dilemmas posed by convergent validation is that every measure must be validated, conceptually in relation to the construct, *in its own right*. I call this validation process *construct-to-measure validation* (see Rossiter, 2013, p. 171). Construct-to-measure, or CtM, validation is a comprehensive form of content validation – a measure design step which, as I pointed out (on pages 308 and 320-322) in my original 2002 article on C-OAR-SE, is conceptualized only superficially in treatises on scale development, such as Nunnally's or Churchill's. CtM validity trumps all other forms of validity including, most importantly, *predictive validity*. As I pointed out in both my 2002 *IJRM* article (pp. 327-328) and my 2011 *EJM* article (p. 1569), predictive validity correlations are completely uninformative because they unjustifiably assume content validity: you cannot know what the true predictive validity correlation is unless you have a perfectly CtM-valid predictor measure in the first place *and* a perfectly CtM-valid criterion measure to relate it to.

My logic-based criticisms of the convergent validation have not been heeded. Social science researchers, and not only those working in marketing research, continue to claim “validity” for new measures of the construct on the basis of convergent correlation (and note that *coefficient alpha*, which is often the only psychometric evidence offered in support of the measure, is merely a form of convergent validation – whereby the measure’s item scores are shown to “converge” on the latent factor or continuum thought to underlie the construct). Well, I now have discovered an indisputable mathematical disproof of convergent validation. This disproof demonstrates, *assuming* a maximally CtM-valid measure has been found, that use of any measure that departs from it – alternative, adapted, shortened, or proxy – will produce not only worse results on average but also will generate results that vary widely over successive studies. If substitute measures cannot be used, the unavoidable conclusion is that for every major construct an optimal measure must be designed. This measure must become the *standard* measure, used in all future research involving the construct.

The task of designing optimal standard measures is thought to be almost impossible in the “soft” sciences such as marketing because so many of our constructs are held to be “abstract” (two similar terms you will come across are “latent” and the ubiquitous catchall term “multidimensional”). But what researchers have overlooked in my C-OAR-SE theory (see my original 2002 *IJRM* article, p. 321; my 2011 *EJM* article, p. 1567; and the 2009 *Transfer: Werbeforschung & Praxis* article by Rossiter and Bergkvist, p. 16) is that all abstract constructs are merely *aggregations of concrete constructs* and that every multiple-item measure – required for all abstract constructs – is made up of *doubly concrete single items*, that is, items which have a clearly comprehended single object coupled with a clearly comprehended single attribute. Thus, at the *questionnaire* level, every measure is a single-item measure, with multiple-item measures simply being an aggregation of single items. The main task in measuring abstract constructs is to specify this aggregation theoretically up front

(instead of selecting a set of items empirically by using factor analysis and coefficient alpha).

This specification constitutes what I call a *construct model*, as introduced in the present article.

The principal purpose of the present article is to explain how construct models can solve the problem of designing optimal standard measures for object-on-attribute *judgment* constructs, the most prevalent type of construct in marketing. (Judgment constructs are “constructed” from three simpler concepts or elements representing an *object*’s rating on an *attribute*, as judged by a particular *rater entity*. These elements are the O, A, and R in the acronym for the C-OAR-SE method.) The next part of the article explains the mathematical dismissal of convergent validation (section 2), a dismissal that provides the necessity to develop optimal standard measures. The last part introduces (in sections 3 and 4) the five new C-OAR-SE-based construct models that will enable the design of optimal standard measures of judgment constructs – multiple-item or single-item, as conceptually required.

2. Mathematical dismissal of convergent validity

This section mathematically dismisses the core psychometric procedure known as “convergent validation.” Researchers readily accept alternative measures of a construct because they believe that a “convergently validated” alternative measure will produce the same findings, but what has not yet been revealed is the great and variable difference in findings that a substitute measure will produce.

2.1 Carlson and Herdman’s revelation and my conclusion

The differences in findings due to different levels of convergent validation were discovered in a recent article in *Organizational Research Methods* by Carlson and Herdman (2012).

Carlson and Herdman in their article do not reject convergent validation. Rather, they

advocate continued use of convergent validation and recommend a convergent correlation minimum of $r = .70$ for accepting an alternative measure of the same construct. Unlike Carlson and Herdman, I conclude that convergent validation must be rejected. My argument goes like this. Consider a 3×3 correlation matrix in which two of the correlations are known, the third not. For example:

	B	C
A	.7	.5
B		?

Most people would guess that the missing correlation r_{BC} would be about .6, the average of the two known correlations. But they would be very wrong. Using a formula developed by Stanley and Wang (1969), and suggested by Yule (1897) in his classic paper on the theory of correlation, in which the average value of the missing correlation is given by $r_{BC} = r_{AB}r_{AC}$, and the limits of the missing correlation are given by plus or minus the square root of the product of $(1-r_{AB}^2)$ and $(1-r_{AC}^2)$, Carlson and Herdman calculated that the missing correlation, r_{BC} , ranges extremely widely in value from $r = -.27$ to $r = .97$, a range which should astound everybody, with a mean value of $r = .35$, which is far lower than anybody would guess. The mean for the missing correlation has to be lower because it is the *product* of the two known correlations, r_{AB} and r_{AC} , rather than the average of them, and the product of two correlations each less than 1.0 will always be lower than either of them. Figure 1 shows this result for the case of a new measure of a predictor construct being substituted for the established measure, the new measure having been accepted because its scores convergently correlate at $r = .70$ with the scores from the established measure. Whereas the established measure of the predictor has a predictive validity coefficient of $r = .50$, the predictive validity coefficient for the new measure of the predictor has a predictive validity coefficient averaging only $r = .35$, and that's an *average*, recognizing that a predictive validity coefficient as low as *negative .27*

or as high as an impossibly redundant .97 could be observed in any single study that uses the new predictor measure.

Figure 1 about here

Now consider a second “missing correlation” example in the following 3×3 matrix:

	C	P
A	.5	?
C		.9

The missing correlation r_{AP} in this example according to Stanley and Wang’s formula, will range from $r = .07$ to $r = .83$, with a mean of $r = .45$. Figure 2 shows this in the case of using a “proxy” measure, P, for the criterion measure, C, selected because scores from the proxy measure correlate very highly with scores from the ideal perfectly valid criterion measure at $r = .90$. The predictor measure, A, remains the same but whereas its true predictive validity coefficient is $r = .50$, it will mispredict the proxy measure of the criterion on average lower, at $r = .45$, and with extreme variation, ranging from .07 to .83, in the possible observed values.

Figure 2 about here

So that researchers can see from their own data just how serious this problem can be, I have prepared Table 1 to show just how far the new mean correlation will drop, and its range expand, depending on the degree of convergent validation accepted to justify the new measure. The top and bottom rows of the table are noteworthy because they represent the extreme cases. As shown in the top row, only in the case of *zero* convergent validity (in other words, orthogonal predictors) will the new predictor’s average predictive validity, which was zero anyway, not change, but the range of the new predictive correlation will be at its very widest. The bottom row shows the purely hypothetical case of *perfect* convergent validity where the new predictive validity correlation does not change, but perfect convergent validity can never be attained because the original predictor measure and the new predictor measure

would have been one and the same. In all other cases, where the convergent validity correlation is *between* zero and 1.0, the predictive validity of the new predictor will always be lower. Also, the lower limit of the new predictor's predictive validity will actually be negative for convergent validity below Carlson and Herdman's recommended minimum of .70 coupled with the original predictor's predictive validity below .70, which is about .20 above what the maximum predictive validity correlation should be for a judgment construct (see Rossiter, 2002, pp. 327-328 and footnote 18 on p. 327). As explained in the note to Table 1, the same figures apply when a proxy criterion measure is substituted for the original criterion measure, as in the second example above.

Table 1 about here

2.2 A different measure will change the findings – usually significantly

Researchers when using a substitute measure are implicitly assuming that the findings produced by it will remain unchanged from those obtained with the original measure.

However, as you can see from Table 1, use of a “convergently validated” substitute measure will, apart from sheer extreme luck whereby the varying new correlation comes out exactly at the *old* value, change the findings. Not only will the same result not be found on average with the new measure (because of the product-of-correlations term that produces the lower mean result) but the results will also be *spread* (normally distributed because of the plus-or-minus term) around the new lower mean correlation. The inevitable spread was demonstrated empirically by Carlson and Herdman (2012) by meta-analyzing numerous studies' findings where a new measure of the same construct was compared for predictive validity with the old. Table 2 shows Carlson and Herdman's meta-analytic results (from their Table 2, p. 27) for three realistic levels of convergent validity: $r = .80$, $r = .70$, and $r = .50$. At $r = .80$, a very high level of convergent validity rarely attained in practice, Carlson and Herdman found that

28% of the 325 pairs of observed predictive validity correlations, which is more than five times the conventional “5% by chance level,” differed in either direction by more than $r = .10$ (which is enough of a difference to change the effect size from Cohen’s “no effect” to “small,” or across the boundary of “small” to “moderate,” or across the boundary of “moderate” to “large,” as well as, of course, enough of a difference to produce boundary-crossing downward changes in effect size). At $r = .70$, which is the minimum degree of convergent validity Carlson and Herdman recommend, 30% of the 345 pairs of predictive validity correlations differed by .10 or more. At $r = .50$, the very low convergent validity level often accepted in the literature, 45% of the observed differences were larger than .10, suggesting not much more than a chance probability of replication of the original finding.

Table 2 about here

2.3 What happens if the original predictor and criterion measures are not CtM-valid?

There is a fundamental assumption in my analysis of convergent validation that goes unrecognized in Carlson and Herdman’s analysis. This assumption is that the original predictor measure, A, and the criterion measure, C, are *perfectly content valid* (in my construct-to-measure, CtM, sense of this term). If this is not the case, then the whole procedure of convergent validation breaks down. Not only does the new correlation between B and C become unstable but so too the original correlation between A and C and the convergent correlation itself between A and B. With three non CtM-valid measures, the 3×3 matrix of measure relationships turns to jelly.

I cannot see how this mathematical dismissal of convergent validation can be argued against. If convergent validation fails as a method of justifying new or substitute measures, the only possible conclusion to be drawn is that we have to find – or more likely develop from scratch – and insist on the use of, in all future research, the optimal CtM-valid measure

of each of our constructs. This conclusion applies to constructs used in every functional role – predictor constructs, moderating or mediating constructs, and criterion constructs.

2.4 *Optimally CtM-valid measures*

Whereas philosophically speaking it is logically possible to derive a *perfect* measure of every construct, the practical reality is that only an *optimal* measure can be achieved. The argument that a proposed measure is optimal in terms of content validity depends on two prior arguments. The first dependency is that the construct has been properly and comprehensively defined (see Rossiter, 2002). Because definitions are always arbitrary and because agreement, being subjective, is also always arbitrary, a faulty or incomplete definition is the first cause of the measure departing from being a perfect representation of the construct. The second dependency is that the construct has been classified correctly, because the classification determines the structure of the measure. C-OAR-SE theory as spelled out in my 2002 *IJRM* article and updated slightly in my 2011 *EJM* article and my 2011 book, to my knowledge, is the only theory of measurement that provides a comprehensive classification system (a classification of object types, attribute types, and rater-entity types). In the present article, I have combined the object and attribute classifications to form the five structural construct models possible for *judgment* constructs.

Section 3 explains the three construct models that require a multiple-item measure, and section 4 explains the two construct models requiring a single-item measure.

3. C-OAR-SE construct models for abstract judgment constructs

Three of the construct models – abbreviated as AC, CA, and AA – involve an abstract object or an abstract attribute, or both. Their possession of even one abstract element means that they require a *multiple-item* measure. An AC construct has an abstract object (requiring

multiple items to represent the main subobjects) measured on a concrete attribute. A CA construct has a concrete object measured on an abstract attribute (requiring multiple items to represent the main subattributes). An AA construct has an abstract object *and* an abstract attribute and is thus “doubly abstract” (requiring multiple items representing the main subobject-subattribute *pairs*). These models are described in the next three sections together with examples of marketing constructs that fit them.

3.1 AC construct model

The AC construct model applies to constructs that are made up of an abstract object coupled with a concrete attribute. The measure becomes a multiple-item measure by virtue of the fact that the abstract object requires separate items to represent the subobjects, while the concrete attribute in the construct remains the same in all items.

Here are three examples of an AC construct. The first is JOB SATISFACTION, where the object is one’s job, the attribute is satisfaction-dissatisfaction, and the rater entity is the job-holder. Most theorists of organizational behavior regard “job” as being an abstract object made up of – that is, formed from – separate subobjects about which the job-holder can be satisfied or dissatisfied, such as the main task that has to be regularly completed, the workload in relation to normal working hours, one’s supervisor (if any), one’s coworkers (if any), and the commute (if any). Assuming that these five subobjects are the major components of a job (and it is easy to check with an open-end-answered survey question whether any major component has been omitted and needs to be added or, conversely, whether any one of the above five is not a major consideration and should be dropped) then each must be represented in a single item, making a 5-item measure in total. The attribute in these items is concrete and always the same and is “satisfaction-dissatisfaction,” with the answer options being a bipolar series of discriminable levels of this attribute.

The second example of an AC construct is the construct called INDIVIDUALISM-COLLECTIVISM, a construct that is becoming increasingly important with the rise of globalism in politics, trade, and of course marketing. “Individualism-collectivism” is actually the *object* in this construct, not the attribute (see the discussion of PERSONAL VALUES as constructs in Rossiter, 2002, pp. 312-313, where it is argued that, in terms of the O-A-R framework for judgment constructs, the “value” is the abstract object, “importance in daily life” is the concrete attribute, and the rater entity is “myself”). As individualism-collectivism is an abstract object, it must be measured by using multiple items; these multiple items represent the most frequently occurring situations in which the person could act in either an individualistic, self-serving manner or else in a collectivistic, group-serving manner. These “go either way” situations, as subobjects, might include support of libertine versus socialistic political candidates, favoring one’s own judgment versus requiring peer agreement for business decisions, and tending to buy self-oriented versus family-oriented household possessions. The attribute in individualism-collectivism is a concrete singular attribute, such as “applies to country X,” in Hofstede’s theory of international individualism-collectivism, (e.g., Hofstede, 1980) or “describes me well,” in Triandis’ theory of personal individualism-collectivism (e.g., Triandis, 1972). Again, the measure is multiple-item only by virtue of the fact that the construct has an abstract object.

The final example of an AC construct is MEDIA PRESENCE, where the object “media” is clearly abstract while the attribute “presence” can be made concrete by reducing it to a count of total length or duration of mentions of the company or person in that medium (see Rossiter, 2013). The abstract object “media” must be represented by one item for each of the major media as subobjects, such as TV, internet, newspapers, magazines, radio, and outdoor.

Given these examples, the structure of the AC model should now make sense. The AC construct model is shown in Figure 3. In the left column are the *concrete* subobjects of the abstract object, with each, being concrete, represented by (the first part of) a single item. In the right column (forming the other part of the item) is the concrete attribute. The judgments required are shown in the middle column and consist of levels of the subobjects' attribute possession – denoted by double subscripts in which the subobject subscript changes but the attribute subscript does not.

Figure 3 about here

3.2 CA construct model

This second type of construct has a structure that is the reverse of the first. A CA construct has a single concrete object that is judged on an abstract attribute made up of concrete subattributes. In this model, multiple items are needed to represent the subattributes. The structure of the CA construct model is shown in Figure 4. In the double subscripts of judged attribute possession in the CA model, the object subscript remains constant while the subattribute subscript changes.

Figure 4 about here

Again, three examples should suffice to illustrate the nature of a CA construct. The first is another organizational behavior construct known as LEADERSHIP POTENTIAL. Here, the concrete object in the construct is a particular candidate for the leadership role whereas the attribute in the construct “leadership potential” is clearly abstract and comprises a number of subattributes. (The rater entity is usually the company’s board of directors or, in the case of political office, members of the candidate’s own political party.) The subattributes indicative of leadership potential could be many (this time best selected from a survey of management experts or political experts, respectively) and might include, as the

main ones, success in lower-level leadership roles, verbal skills such as clear communication, and a record of consistently astute decision-making.

The second example of a CA construct is CORPORATE IMAGE, which is generally defined as the set of immediate cognitive and affective associations (the subattributes) that come to mind (and here the rater entity could be prospective, current, or past customers; or, a different construct by dint of a different rater entity, prospective, current, or past employees) when one thinks of the company as a whole (the single concrete object). Again, the items are multiple only over the subattributes. The most common mistake in designing a measure of corporate image is to force on respondents a set of “image attributes” selected from a larger pool of items by factor analysis. There is no good reason why the image attributes should form “factors” – they don’t in the respondents’ minds, so why should they in the researcher’s mind – and, most important for academic market researchers to realize, managers cannot operate with abstract factors when trying to improve or change the corporation’s image.

The third CA construct example is STORE ATMOSPHERE. The object – the store or other retail venue – is concrete, but the attribute of “atmosphere” is theorized as abstract and is composed of two separate subattributes called pleasantness and arousal. In Donovan and Rossiter’s (1982) well-known study testing this theory, the usual practice was followed of employing multiple items (taken from Mehrabian and Russell, 1974) to measure the respondent’s level of arousal and the respondent’s level of positive or negative affect. However, according to Rossiter (2002), these two subattributes are concrete, so single-item measures of both should have been used, with a gradated unipolar answer scale for arousal and a gradated bipolar answer scale for affect. The scoring of the two subattribute ratings, however, would remain as in the original because in Donovan and Rossiter’s store-atmosphere theory, which follows Hull’s (e.g., 1952) Drive $\times$ Habit-Strength theory, the two emotional states of unipolar arousal and bipolar affect *multiply*; in short, the manager should

add excitational in-store stimuli to raise the typical shopper's arousal level if the store is pleasant to begin with, but dampen the stimuli to lower the arousal level if shoppers' initial impression of the store is unpleasant. The multiplicative scoring rule, of course, goes against the usual practice of adding (and then averaging) items' ratings.

3.3 AA construct model

The most complex construct model underlying judgment constructs is the AA, or “doubly abstract,” construct model, where both the object and the attribute of the construct are abstract in that they carry multiple major meanings. The object consists of subobjects and the attribute consists of subattributes, and the items form subobject-subattribute *pairs*, with a subobject often paired with multiple subattributes. The AA construct model is shown in Figure 5. There are now triple subscripts in the AA construct model to accommodate the complexity of item composition.

Figure 5 about here

Inability of measure designers to realize that the items are subobject-subattribute pairs leads to the biggest problem in AA construct measure design – when psychometrically trained researchers, as they inevitably do, factor-analyze the multiple items' scores. What happens is that the “factors” that emerge turn out to be almost uninterpretable, even by academics let alone managers, because the correlations between the items' scores that produce the “factor” are based on respondents variously perceiving the items as related by virtue of *either* their common subobject *or* their similar subattributes. This is undoubtedly what happened with Parasuraman, Zeithaml, and Berry's (1988) SERVQUAL measure of COMPANY SERVICE QUALITY. In the 22-item SERVQUAL measure there are three different subobjects – “the company” (9 items), “physical facilities” (3 items), and “employees” (10 items) – which are paired with miscellaneous subattributes. In the

SERVQUAL questionnaire, the *subobject* is the response-grouping stimulus most salient to the respondents. The researchers, on the other hand, factor analyzed the respondents' ratings and grouped them only on the basis of *subattributes* – hence the factors were labeled as “reliability,” “assurance,” “responsiveness,” and the like. How, instead, should the service quality measure have been designed? The subobjects that are the “agents” for that type of company's service-quality subattributes should first be identified from qualitative research, together with only the main subattributes that constitute “service” in relation to the particular subobject (Rossiter, 2009). These subobject-subattribute pairs become the items and must not be subjected to a factor analysis to delete any of them. In the interest of illustrating how this works, the present author offers here an efficient (12 items answered “yes” or “no”) generic measure of company service quality that conforms to the AA construct model. This efficient, AA-conforming questionnaire is presented in Table 3. The content of this measure can be debated, but not its model-based structure.

Table 3 about here

Another major construct in marketing whose measure should be based on the AA construct model is the company's MARKET ORIENTATION. Market orientation (to call it by its brief attribute-only name) is a complicated construct that has been conflictingly defined and differentially poorly measured by academic market researchers. There is insufficient space to go into the conceptual problems here but interested readers might consult Rossiter's (2012a) commentary article which criticizes the leading measures and outlines a more content-valid approach.

4. C-OAR-SE construct models for concrete judgment constructs

The remaining two construct models for judgment constructs – abbreviated as CC and CDSA respectively – require only a single-item measure. In fact, as will be demonstrated, the

measure is *invalidated* if a multiple-item measure is used. The first, CC, is the “doubly concrete” construct model (single concrete object coupled with a single concrete attribute). The second, CDSA, refers to a special type of construct that combines a concrete object with two concrete subattributes. The CC and CDSA construct models are described in the next two sections.

4.1 CC construct model

The CC (“doubly concrete”) construct model is the basic model for judgment constructs where the object is concrete (single meaning) and the attribute on which it is being judged is also concrete (single meaning). Figure 6 shows the CC construct model.

Figure 6 about here

There are many consumer behavior examples of constructs that fit the CC model. Those exemplified in Bergkvist and Rossiter’s (2007) “single-item vs. multiple-item” *JMR* article, for example, were AD LIKING, BRAND BENEFIT BELIEF(S), BRAND ATTITUDE, and BRAND PURCHASE INTENTION.

The biggest mistake that researchers make in designing measures of doubly concrete constructs is to use multiple items in the belief that they will cover “various facets of the construct” (and here note that researchers invariably mean various facets of the *attribute*, neglecting the construct’s object altogether). They do this most often by choosing the main meaning of the construct to become the first single item and then adding further items to cover less common but possible meanings. Multiple-item theorists argue that the additional items cannot do any harm to the measure because they are fairly synonymous, though not completely redundant, with the first item. The problem with this argument is that every additional item will necessarily be less synonymous with the first item (as the many researchers who have had to dream up additional items to help boost coefficient alpha would

know) but that their scores are *weighted equally* with scores from the first, would-be, single item. This faulty practice will always move individuals' multiple-item observed scores away from their single-item true scores. An example from Osgood, Suci, and Tannenbaum (1957) will demonstrate why this is so. Table 4 lists the factor loadings (taken from Osgood et al., pp. 53-61) of various adjective-pairs used to “cover different facets of” the construct called EVALUATION in the Evaluation-Potency-Activity theory of connotative meaning. The (paired antonymous) adjective ratings were taken over a diverse set of objects and so the ratings are highly generalizable. As can be seen, the “marker item” for Evaluation is “good-bad” with a factor loading set to 1.00, and the other adjective pairs have successively lower loadings, ranging from .68 down to .40. In a multiple-item measure of the Evaluation construct, all these items would be given equal weight no matter how “off attribute” they are. Perversely, the only reason why multiple-item scores might turn out to be any less “off attribute” would be, as Drolet and Morrison (2001) alleged, common-measure response bias, whereby many respondents “straight-line” their answers based on their answer to the first item.

Table 4 about here

4.2 CDSA construct model

This section of the article introduces the construct model for designing a construct-to-measure valid single-item measure of a construct that is concrete in the object but abstract in the attribute in that the attribute is composed of two concrete and necessary subattributes. This model is called the CDSA construct model, with the label an acronym for “concrete object, dual subattribute.” The CDSA construct model is shown in Figure 7. As shown, the single item is made up of a concrete object representation coupled with two jointly necessary subattributes, called subattribute 1 and subattribute 2 in the diagram. The judgment required

is a binary one – “yes” or “no” – and in answering “yes” the rater is attesting that both subattributes are present and in answering “no” the rater is attesting that at least one subattribute is absent.

Figure 7 about here

It is important to note that a CDSA measure is deliberately “double-barreled” but not in the usual erroneous sense of asking the rater to make a conditional instead of an unconditional judgment. (An example of a conditionally double-barreled question is “Are you in favor of building more nuclear power plants so that we can have enough electricity to meet the country’s needs?”; see Bradburn, Sudman, and Wansink, 2004, p. 143.) CDSA measures are different in that they asked the rater to make a *conjoint* judgment about two of the object’s *subattributes*. Asking the rater to make separate judgments about the two subattributes, as separate items, is not a legitimate option – for the simple reason that the summing or averaging of the ratings will conceal the fact that positive answers need to be given to both subattributes. The fallacy in the separate items approach is easily realized in the following examples.

The CDSA construct model was first applied in the study by Rossiter (2012b) to design a single-item measure of the construct of *quasi-romantic brand love*. The construct’s object, the branded product, is concrete, and from a search of the psychological and sociological literature on human romantic love Rossiter identified and argued for defining quasi-romantic brand love in terms of two necessary and sufficient subattributes, which he called “deep affection” and “separation anxiety.” Rossiter’s CDSA-based, single-item measure of quasi-romantic brand love is as follows:

I would say I feel deep affection, like love, for this brand and I
would be really upset if I couldn’t have it
☐ Yes ☐ No

The two defining subattributes, worded to represent *quasi*-love as “deep affection, like love” and to represent separation anxiety in ego-preserving terms as “I would be really upset if I couldn’t have it,” are being *jointly* attested to when the consumer gives an answer of “yes.” On the other hand, when the consumer gives an answer of “no,” the logical implication is that one or both of these subattributes is lacking. Also worthy of note is that this measure’s answer options are “forced” binary in that the rater has to answer either yes or no, whereas in an earlier study Rossiter and Bellman (2012) employed the same item with a yes-only, or “pick any,” answer option whose “unforced” nature, as demonstrated by Dolnicar, Rossiter, and Grün (2012), makes answering the measure unstable on repeated measurements such as used in brand tracking studies.

Two further examples of CDSA constructs are the emotional states of *happy* and *sad*. Researchers in clinical psychology and also in consumer behavior typically have measured the *happy* emotional state on a continuous answer scale, which means that researchers accepted any rating from “slightly happy” through to “extremely happy” as all representing the same state that we call happiness. This is in conflict with the Webster’s Dictionary definition of happiness, which states that this emotional state is composed of two simultaneously present subattributes: “pleased” and “contented.” (Again, remember that we are trying to measure whether or not the individual feels happy in our defined sense; we are not, as most researchers believe they have to do, trying to find out how the proverbial “person in the street” defines happiness.) These two subattributes cannot be measured as separate items, nor can the single item “I am feeling happy” be used, because as a single item the attribute “happy” remains abstract and undefined. The required measure is a dual-subattribute single item with a binary answer, thus:

At present, I am feeling pleased and contented.

☐ Yes ☐ No

A “yes” answer definitionally consistently signifies the presence of the emotional state called *happy*.

And how do you measure *sad*? Forgas (2013), a leading theorist of emotions, is careful to distinguish the state of *sadness*, as a mild and temporary level of negative affect, from “intense, enduring, or debilitating *dysphoria*” (p. 230, emphasis added here) symptomatic of Major Depressive Disorder. A valid measure of *sadness* – according to Forgas’s definition of the construct – would therefore be a CDSA single-item measure along the lines of:

At present, I feel sad, though not so much that I am depressed.

☐ Yes ☐ No

In this CDSA measure, the two subattributes that together denote the emotion of sadness, as distinct from the clinical state of depression, are: “sad” and “not depressed.”

5. Where to from here?

Researchers in the social sciences have an enormous task ahead of them if they are to transform their particular fields into legitimate scientific enterprises based on sound measurement (Rossiter, 2011b). The need for a total and radical transformation of measure design is undeniable because of the failure of psychometrics as demonstrated in the present article by the dismemberment of its core assumption, convergent validity. Rossiter’s C-OAR-SE method, especially as enhanced by the addition of the five construct models, appears to be a comprehensive and practical replacement for the conventional psychometric approach to the design of measures. Social science researchers, marketing researchers among them, are going to have to cease regarding C-OAR-SE as an alternative and seemingly too difficult measure-design method and get to terms with it as a replacement for the psychometric method.

Almost all the constructs in the social sciences are either *recognition* constructs, *recall* constructs, or – the most prevalent type as examined in the present article – *judgment* constructs. While most measures of recognition and recall constructs are very badly designed – see especially Rossiter and Percy (1997) for various valid and non-valid measures of "brand awareness," a construct that involves either recognition or recall and in one circumstance both – the most extensive threat to our marketing knowledge stems from the mismeasurement of *judgment* constructs. Here is what must be done. First, researchers must define each judgment construct "tripartite" in terms of the focal object and its subobjects if any, the focal attribute and its subattributes if any, and then must specify the rater entity who will be performing the judgment of the object in terms of the attribute. Second, researchers must decide which (judgment) construct model the construct corresponds to – one of the five given in this article. Third, researchers must then write items that optimally fit the object and attribute elements of the construct as specified by its definition and its construct model, and also write answer options that fit the language usage and discrimination capacity of the rater entity. Last, a thorough "cognitive interview" pretest – see Bolton (1993), Conrad and Blair (2009), and Blair and Conrad (2011) for guidance on this – must be conducted with a reasonable sample of the least-educated target raters to ensure that the wording of items and answer options is as clear as possible so that none of the raters should make "errors" when recording their judgments. The measure is then ready to use. The measure does not require any post-use "backwards justification" by appealing to the psychometric statistics it produces, which are anyway totally unnecessary.

Having developed an optimal construct-fitting measure for the research, the *really* difficult task then begins, and this is the task of persuading other researchers in the same field to use this measure as the *standard* measure of the construct. To accomplish this task, researchers are going to have to accept the arguments and evidence presented in this article

demonstrating that different, even slightly different, measures of the same construct will produce different findings. When it comes to measurement, “near enough is *not* good enough.” Marketing, like other social fields, has no hope of becoming a science until researchers develop and agree on an optimal standard measure of each of its main constructs.

7. References

- Bergkvist, L., & Rossiter, J.R. (2007). The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of Marketing Research*, 44(2), 175-184.
- Blair, J. & Conrad, F.G. (2011). Sample size for cognitive interview pretesting. *Public Opinion Quarterly*, 75(4), 636-658.
- Bolton, R. N. (1993). Pretesting questionnaires: content analyses of respondents' concurrent verbal protocols. *Marketing Science*, 12(3), 280-303.
- Bradburn, N.M., Sudman, S., & Wansink, B. (2004). *Asking Questions: The Definitive Guide to Questionnaire Design*. San Francisco: Jossey-Bass.
- Campbell, D.I., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81-105.
- Carlson, K.D., & 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.
- Conrad, F.G., & Blair, J. (2009). Sources of error in cognitive interviews. *Public Opinion Quarterly*, 73(1), 32-55.
- Cronbach, L.J., & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281-302.
- Dolnicar, S., Rossiter, J.R., & Grün, B. (2012). 'Pick any' measures contaminate brand image studies. *International Journal of Market Research*, 54(6), 821-834.
- Donovan, R.J., & Rossiter, J.R. (1982). Store atmosphere: an environmental psychology approach. *Journal of Retailing*, 58(1), 34-57.

- Drolet, A., & Morrison, D. (2001). Do we really need multiple-item measures in service research? *Journal of Service Research*, 3(3), 196-204.
- Forgas, J.P. (2013). Don't worry, be sad! On the cognitive, motivational and interpersonal benefits of negative mood. *Current Directions in Psychological Science*, 22(3), 225-232.
- Hofstede, G. (1980). *Culture's Consequences: International Differences in Work-Related Values*. Beverly Hills, CA: Sage.
- Hull, C.L. (1952). *A Behavior System*. New Haven: Yale University Press.
- Mehrabian, A., & Russell, J. (1974). *An Approach to Environmental Psychology*. Cambridge, Mass: MIT Press.
- Nunnally, J.C. (1978). *Psychometric Theory* (2nd edn.). New York: McGraw-Hill.
- Osgood, C.E., Suci, G.J., & Tannenbaum, P.H. (1957). *The Measurement of Meaning*. Urbana, IL: University of Illinois Press.
- Parasuraman, A., Zeithaml, V.A., & Berry, L.L. (1988). SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, 64(1), 12-40.
- Rossiter, J. (2002). The C-OAR-SE procedure for scale development in marketing. *International Journal of Research in Marketing*, 19(4), 305-335.
- Rossiter, J. (2009). ER-SERVCOMPSQUAL: a measure of e-retailing service components quality. *Service Science*, 1(4), 212-224.
- Rossiter, J.R. (2011a). Marketing measurement revolution: the C-OAR-SE method and why it must replace psychometrics. *European Journal of Marketing*, 45(11/12), 1561-1588.
- Rossiter, J.R. (2011b). *Measurement for the Social Sciences: The C-OAR-SE Method and Why It Must Replace Psychometrics*. New York: Springer.

- Rossiter, J.R. (2012a). Further comment on “market orientation.” *Australasian Marketing Journal*, 20(1), 108-112.
- Rossiter, J.R. (2012b). A new C-OAR-SE-based content-valid and predictively valid measure that distinguishes brand love from brand liking. *Marketing Letters*, 23(2), 905-916.
- Rossiter, J.R. (2013). Scientific progress in measurement theory? *AMS Review*, 3(3), 171-179
- Rossiter, J.R., & Bellman, S. (2012). Emotional branding pays off. *Journal of Advertising Research*, 52(3), 291-296.
- Rossiter, J.R., & Bergkvist, L. (2009). The importance of choosing one good item for single-item measures of attitude towards the ad and attitude towards the brand and its generalization to all measures. *Transfer Werbeforschung & Praxis*, 55(2), 8-18.
- Stanley, J., & Wang, M. (1969). Restrictions on the possible value of r_{12} , given r_{13} and r_{23} . *Educational and Psychological Measurement*, 29(3), 579-581.
- Triandis, H.C. (1972). *The Analysis of Subjective Culture*. New York: Wiley.
- Yule, G.U. (1897). On the theory of correlation. *Journal of the Royal Statistical Society*, 60(4), 812-854.

Figure 1

Range of predictive validity coefficients observable with a “convergently valid” new predictor measure

Figure 2

Range of predictive validity coefficients observable with a “convergently valid” proxy measure of the criterion

Figure 3

AC construct model: abstract object, concrete attribute

Figure 4CA construct model: concrete object, abstract attribute

Figure 5

AA construct model: “doubly abstract”

Figure 6

CC construct model: “doubly concrete”

Figure 7

CDSA construct model: concrete object, dual subattribute

Table 1: Mean (minimum, maximum) values of the missing correlation, r_{BC} , when r_{AB} and r_{AC} are known. Calculated from the formula $r_{BC} = r_{AB} r_{AC} \pm \sqrt{(1 - r_{AB}^2)(1 - r_{AC}^2)}$

Convergent correlation of predictor A and predictor B (r_{AB})	Predictor A correlation with criterion C (r_{AC})					
	.30	.40	.50	.60	.70	.80
Orthogonal .00	.00 (−.95, .95)	.00 (−.90, .90)	.00 (−.87, .87)	.00 (−.80, .80)	.00 (−.71, .71)	.00 (−.60, .60)
.30	.09 (−.82, 1.00)	.12 (−.75, .99)	.15 (−.68, .98)	.18 (−.54, .90)	.21 (.47, .89)	.24 (−.33, .81)
.40	.12 (−.75, .99)	.16 (−.68, 1.00)	.20 (−.59, .99)	.24 (−.49, .97)	.28 (−.37, .93)	.32 (−.23, .87)
.50	.15 (−.68, .98)	.20 (−.59, .99)	.25 (−.50, 1.00)	.30 (−.39, .99)	.35 (−.27, .97)	.40 (−.12, .92)
.60	.18 (−.54, .90)	.24 (−.49, .97)	.30 (−.39, .99)	.36 (−.28, 1.00)	.42 (−.15, .99)	.48 (.05, .91)
.70	.21 (−.47, .89)	.28 (−.37, .93)	.35 (−.27, .97)	.42 (−.15, .99)	.49 (−.02, 1.00)	.56 (.13, .99)
.80	.24 (−.23, .87)	.32 (−.23, .87)	.40 (−.12, .92)	.48 (.05, .91)	.56 (.13, .99)	.64 (.28, 1.00)
.90	.27 (−.15, .69)	.36 (−.04, .76)	.45 (.07, .83)	.54 (.19, .89)	.63 (.32, .94)	.72 (.46, .98)
Perfect 1.00	.30 (.30, .30)	.40 (.40, .40)	.50 (.50, .50)	.60 (.60, .60)	.70 (.70, .70)	.80 (.80, .80)

*This table can be alternatively used to read the mean, minimum, and maximum values of the predictive validity of a predictor for predicting a proxy criterion, r_{PC2} , by substituting P = predictor for A, the proxy C2 for B, and C1 for C. The convergent correlation between (scores on) the original criterion and the proxy criterion is then r_{C1C2} , equivalent to r_{AB} above. Even a very good proxy measure that correlates at $r = .90$ with the original criterion measure will necessarily be predicted worse, on average, due to the new predictive validity correlation being .90 times the original. For example, if the original predictive validity, r_{PC1} , is .50, then the new predictive validity, r_{PC2} , will average $.90 \times .50 = .45$, with a possible minimum in any single study of .07 and possible maximum of .82.

Table 2

Meta-analysis of 300+ observed differences in predictive validity correlations for the original measure and the new measure convergently correlating at three levels (results extracted from Carlson and Herdman, 2012, Table 2)

Convergent correlation	N	Difference magnitude						Percent greater than $\Delta r = .10$
		Percent .11 – .20		Percent .21 – .30		Percent >.30		
.80	(325)	22	+	3	+	3	=	28 %
.70	(345)	21	+	6	+	3	=	30 %
.50	(377)	27	+	11	+	7	=	45 %

Table 3

Typical items for measuring componential service quality according to the AA construct model

Subobjects	Paired subattributes
1. Typical member of sales staff	1a. Well-informed 1b. Helpful 1c. Courteous
2. Retail stores (if company has them)	2a. Good product range 2b. Can quickly find what you're looking for
3. Retail website (if company has one)	3a. (same as 2a) 3b. (same as 2b) 3c. Simple to navigate 3d. Easy to change order 3e. Good speed vs. cost delivery options
4. Returns	4a. Liberal return policy 4b. Convenient to make return 4c. Acceptable refund/credit policy

Table 4

Evaluative items are not equal: correlations – factor loadings – of evaluative adjective pairs on the Evaluation factor (from Osgood et al., 1957)

Adjective pair	Evaluation factor loading
Good-bad	1.00
Reputable-disreputable	.68
Wise-foolish	.57
Beautiful-ugly	.52
Kind-cruel	.52
Successful-unsuccessful	.51
True-false	.50
Positive-negative	.48
Interesting-boring	.40