The current issue and full text archive of this journal is available at
www.emeraldinsight.com/0309-0566.htm

COMMENTARY

Avoiding measurement dogma: a response to Rossiter

Edward E. Rigdon
Georgia State University, Atlanta,Georgia, USA
Kristopher J. Preacher
University of Kansas, Lawrence, Kansas, USA
Nick Lee
Aston University, Birmingham, UK
Roy D. Howell
Texas Tech University, Lubbock, Texas, USA
George R. Franke
University of Alabama, Tuscaloosa, Alabama, USA, and
Denny Borsboom
University of Amsterdam, Amsterdam, The Netherlands

Abstract

Purpose – This paper aims to respond to John Rossiter's call for a "Marketing measurement revolution" in the current issue of *EJM*, as well as providing broader comment on Rossiter's C-OAR-SE framework, and measurement practice in marketing in general.

Design/methodology/approach – The paper is purely theoretical, based on interpretation of measurement theory.

Findings – The authors find that much of Rossiter's diagnosis of the problems facing measurement practice in marketing and social science is highly relevant. However, the authors find themselves opposed to the revolution advocated by Rossiter.

Research limitations/implications – The paper presents a comment based on interpretation of measurement theory and observation of practices in marketing and social science. As such, the interpretation is itself open to disagreement.

Practical implications – There are implications for those outside academia who wish to use measures derived from academic work as well as to derive their own measures of key marketing and other social variables.

Originality/value – This paper is one of the few to explicitly respond to the C-OAR-SE framework proposed by Rossiter, and presents a number of points critical to good measurement theory and practice, which appear to remain underdeveloped in marketing and social science.

Keywords Classical measurement theory, Psychometric tests, Measure validity, Marketing, Measurement

Paper type Viewpoint

We share with Rossiter the concern stated in his call for a "Marketing measurement revolution" (a paper hereafter referred to as MMR) in this issue of *EJM*, and in his prior work (e.g. Rossiter, 2002, 2008, 2011), that knowledge development in marketing and

European Journal of Marketing
Vol. 45 No. 11/12, 2011
pp. 1589-1600

0309-0566
DOI 10.1108/03090561111167306

the social sciences more generally is hampered by suboptimal measurement practices (e.g. Borsboom, 2006; Franke *et al.*, 2008; Howell *et al.*, 2007; Lee and Hooley, 2005; Rigdon, 1994). In our view, one of the main reasons for this is the ritualistic adherence to dogma concerning the conceptual and empirical analysis of test scores. We hope that MMR (and this response) will provoke a much-needed discussion on these topics. However, while we applaud the critical spirit of MMR, and at least partly agree on Rossiter's diagnosis of the pathology, we disagree with the suggested cure. In fact, MMR's apparent panacea – the blanket rejection of all of what MMR terms "psychometric" techniques currently in existence, and the wholesale adoption of the C-OAR-SE method – would seem to perpetuate the problem. More specifically, the recommendation looks to replace one dogma with another, rather than lead to an enlightened approach to measurement. Instead, we entreat researchers to think far more broadly about the range of possible relationships between observable measurements (e.g. item or test scores, behaviors, and suchlike) and the attributes being measured, and to employ measurement models that are accurate representations of these relationships as far as possible, rather than accept on faith that any single model applies in all cases.

In the following comment, we respond to MMR along three main lines of argument. We will begin by explaining what we see as the major faults in marketing measurement. The reader will find that many of our concerns overlap with those discussed in MMR. However, subsequently, we will show that our own views on the best response to these problems diverge greatly from those presented in MMR, which we will argue to be flawed in at least three key ways. Finally, we will suggest a number of recommendations for how marketing science (and by extension social science in general) can improve current measurement theory and practice.

What is wrong with measurement practice in marketing?

Ritualistic measurement practice

We agree with Rossiter and others that marketing and social science research is rife with the uncritical application of existing measurement procedures. However, we would argue that in such fields, there is also a tendency towards an unthinking acceptance of so-called "new" measurement techniques. Thus, an overarching picture of research practice would undoubtedly show an almost reactionary tendency of researchers to move *en masse* from one method to another, without in-depth questioning of the appropriateness of any new method. Thus, we welcome Rossiter's work, as well as the formative measurement literature for example, only if it leads researchers to think more carefully about measurement operations, rather than to uncritically adopt whatever new method is proposed.

Much current thinking about the relationship between observables and theoretical constructs undeniably proceeds along the "default" lines of standard psychometric theory. That is, many researchers automatically assume that the relationship between observables and constructs is reflective, such that multiple observables reflect the common influence of a latent variable. Other ways of conceptualizing such relationships exist, however; for instance, the relation may be one in which the observables; determine the construct (the formative model from Edwards and Bagozzi in 2000); are samples from a behavioral domain (McDonald, 2003); depend on partly

overlapping cognitive processes (Bartholomew *et al.*, 2009); or, are directly related elements of a causal system (Borsboom, 2008; Van der Maas *et al.*, 2006; Cramer *et al.*, 2010). Each of these examples describes a legitimate relation between observables and theoretical constructs that is not incorporated in the reflective model.

Certainly, the relationship between the observables and the construct may be such that the researcher is free to define his or her construct solely on the basis of theoretical principles or expert opinion, as recommended in MMR. In these cases, the C-OAR-SE scheme may offer a useful guide for measure development. However, just as it is dogmatic to assume that in every situation a reflective model is appropriate, imposing the idea that the C-OAR-SE model is instead always appropriate is simply replacing one dogma with another.

Over-reliance on statistical models at the expense of construct definition

In MMR we hear echoes of Pearl's (2010) criticism of the impact of statisticians on structural equation modeling, which, Pearl agues, reduced a valuable technique for understanding causal relations into a mere generalization of regression. However, where Pearl argues that careful logical analysis should lead to empirically testable counterfactual statements, MMR instead states that all empirical evidence is irrelevant to the evaluation of measures.

We agree that content considerations are seriously undervalued in contemporary measure development. In fact, the very concern for content coverage is inconsistent with the reflective model. That is, in the reflective model, items are exchangeable apart from measurement characteristics such as difficulty and precision. In other words, given a subject's true score on one item for a unidimensional reflective model, there is no need to use any other items. For example, consider a set of thermometers which are truly unidimensional and reflective of ambient temperature (save for random error). If one knows the expected value (long-run average under identical conditions) of one thermometer then one does not have to look at another. Conversely, consider a typical set of psychometric items – say for the construct of depression. Here, if one knows the subject's expected value on the item "do you sleep well at night," one remains interested in the response to an item such as "do you plan on committing suicide". Of course, these items are empirically correlated, and perhaps causally related, but they are not actually exchangeable in the way the reflective model assumes they are. Thus, their treatment as reflective measures can produce misleading results. Rossiter would presumably agree with our contention that many marketing and social science item sets fall into the second category, and the default reliance on a reflective scheme can therefore be problematic.

The failure to recognize this issue can be seen in much modern measurement practice, which often leads to the use of a few highly-correlated indicators to measure thin slices of constructs. However, in creating such measures, researchers may have removed a high percentage of the original item set in order to achieve empirical unidimensionality and high reliability. Thus, these researchers are often no longer testing their original theory since they measure but a single part of a more general construct. If so, they should then return to their original theory to make certain their measures are consistent with their conceptualization. MMR is correct in showing how such practices can lead to poor theory testing. Yet such an iterative process is explicit

in expressions of the psychometric model by authors such as Churchill (1979). It is not the model, but the application, which is problematic (a point to which we will return).

The opposite problem is also evident, when researchers start with conceptually complex constructs that encompass distinct but related subdimensions. Measuring such complex constructs leads to either aggregating multiple concepts into a single scale, consequently reducing interpretability, or to creating multidimensional scales, consequently complicating measure development and validation. Such complex constructs are often the result of either a lack of conceptual clarity in their original definition, or a convenient verbal label for a set of items in the same general conceptual domain that is mistaken for a construct to be measured. For example, while variables in the socioeconomic status (SES) domain may include such things as education, income, housing status, and occupational prestige, these elements do not make SES a real entity. Certainly, these elements may be correlated, but their correlation is not because a construct of SES is their common cause. The formative literature argues that the formative model is therefore appropriate, but this model still appears to treat SES as a unitary entity. While this may be convenient for communication, it may also result in the loss of valuable information on the unique relationships between the items and their outcomes (e.g. Franke *et al.*, 2008). In fact, studies that have compared scales (whether aggregate, higher-order factors, or formative) composed of conceptually distinct constructs at different levels of complexity have consistently found that both prediction and understanding are enhanced by using a larger number of specific variables rather than a smaller number of more global ones (Mershon and Gorsuch, 1988; Paunonen, 1998; Schneider *et al.*, 1996; Edwards, 2001; Howell *et al.*, 2007; McGrath, 2005, 2009).

A limited consideration set of measurement models

It is clear then, that we agree with MMR that the near-exclusive reliance of marketing and social science research on the reflective model is harmful. In fact, we would also say that the apparent situation in contemporary marketing science – where if something is "not reflective", then it must be "formative" – is almost as harmful. However, we stand diametrically opposed to the demands espoused in MMR that essentially all measurement must now conform to a single paradigm.

Instead, we encourage researchers to consider the broad range of possible relations between measures and constructs. This range is limited only by our ability to imagine and implement measurement models. Computational advances over the years have dramatically expanded this potential range, but the scope of our imagination broadens more slowly, as newly-acknowledged possibilities allow us to imagine just a little further. At present, the many possible measurement forms might be arranged along at least five dimensions: linear vs. nonlinear; single vs. multiple; factor vs. component; direct vs. indirect; and strict vs. relaxed. While not necessarily the final word in measurement taxonomies, these five dimensions already map a wider space for measurement models than marketing researchers have been inclined to explore.

Relations between measures and constructs may be linear or nonlinear. The standard confirmatory factor model, and the approach of MMR, are both examples of linear relations. However, methods are available for modeling nonlinear relations, even

within factor models (Muthén and Muthén, 2010), and nonlinear logit or probit forms may make sense when either measure or construct is discrete (Reise *et al.*, 1993).

Single or multiple measures may be associated with single or multiple constructs. Here, Rossiter's approach lies strictly within the single/single "cell." Factor model researchers have long recognized the possibility of observed variables loading on multiple constructs (Rindskopf and Rose, 1988; Mulaik and Quartetti, 1997). Partial least squares path modeling is generally employed to relate sets of measures to single component-based constructs, but early work in this field envisioned the extraction of multiple components from a given set of measures (Apel and Wold, 1982).

Researchers might structure their constructs as factors/latent variables or as component-like composites. It may be that the factor/component dimension subsumes the reflective/formative distinction, which is so well known among academic researchers. Here, structural equation modeling and partial least squares path modeling stand as ready examples of the two alternatives.

Measurement methods may involve either more direct or more indirect connections between the measurement and the attribute being measured. Measurement is direct to the extent that it focuses on immediate consequence of the attribute itself. Ambient temperature gives rise to thermometer readings almost directly, and retail sales give rise to point-of-purchase sales records in a similar way. MMR asserts that direct methods are applicable and preferable for a broad range of marketing constructs. Indirect measurement removes the observation from the immediate vicinity of the attribute. Conjoint measurement and varieties of unidimensional and multidimensional scaling allow researchers to derive evaluations of product attributes implicitly. Direct measurement shortens the chain of inferences necessary to establish measurement validity, but more indirect measurement may allow researchers to sidestep specific threats to validity such as social desirability bias, and may sometimes be the only method available. Of course, closer examination of any measurement procedure will reveal mediating mechanisms and potential confounds in even the seemingly most direct of approaches.

Measurement models may include either strict, tightly constrained relations or may specify relations that are more flexible but less precise. Historically, researchers have favored strict models, even when those models must be subsequently relaxed, as is often the case with confirmatory factor models. MMR's proposal certainly falls at the "strict" end of this dimension. Instead of asserting strict models and then engaging in risky ex-post modification (Cliff, 1983), researchers could acknowledge the limitations of their prior knowledge by adopting relaxed models, whether factor models with additional parameters (Asparouhov and Muthén, 2010) or the set correlation method proposed by Cohen (1982). In Asparouhov and Muthén's (2010) approach to Bayesian confirmatory factor analysis, the strictly-zero cross-loadings of conventional CFA are replaced by parameters with near-zero expected values and small variances, allowing for CFA models which do not impose a strict simple structure but which perhaps better represent actual measurement phenomena. Cohen's set correlation method captures the full linear effect of sets of independent and dependent variables, without specifying (beyond mere inclusion) the relationships between the observed variables and individual linear composites.

As we see it, researchers have many choices, with consequences for the inferences that they may subsequently draw regarding their constructs. Researchers who make unconventional choices may be challenged regarding their inferences, and indeed may be uncertain regarding what sorts of inferences are justified. Then again, the marketing research community has never fully confronted the impact of ex-post modifications, ranging from the relaxation of specific constraints (e.g. correlating error variances) to the discarding of items, which have long accompanied conventional measurement practices. There is much work to be done. Retreating into dogma may seem like a way to avoid addressing these difficult questions, but ignoring these questions undermines the quality of research, raising questions about the value of the entire research enterprise.

Our current situation, where the universe of measurement models includes only the reflective and formative perspectives, leads to a major dilemma when a measurement model fails its tests. By contrast, a world with multiple alternative measurement models requires researchers to give serious *a priori* consideration to their choice of measurement model rather than rely on tradition and dogma. Indeed, researchers may then plausibly consider alternative or fall back measurement models if their first choice model fails. Surely, this situation must be an improvement over our current practice. However, it is also an improvement on the future outlined in MMR, where we replace one methodological straitjacket with another.

What is wrong with MMR's arguments?

Validity assertions are immune to empirical challenge

The C-OAR-SE approach relies heavily on expert judgment as its central pillar, and as a point of departure from existing approaches to measurement. Under C-OAR-SE, a measure is valid if a self-proclaimed measurement expert says it is. What to do if experts disagree then? While existing statistical and/or psychometric methods at least provide some degree of objectivity, the total reliance of C-OAR-SE on expert opinion seems to place measurement on the same pseudo-scientific plane as phrenology (Chaline, 2009). It also reminds the authors of a certain cheesemaker, who experienced a string of failed attempts to make varieties of cheese she had made successfully in the past. Cheesemakers use microorganisms (and enzymes) to turn milk into cheese curd. Different microorganisms require the milk to be at different temperatures. The wrong temperature may retard the transformation or even kill the microorganisms. This particular cheesemaker (clearly an expert in MMR's terms) measured milk temperature (in MMR's terms a "hard" measure) with a thermometer, which she had every reason to believe that it was valid. However, our cheesemaker friend learned otherwise when she dropped the thermometer into boiling water and observed the reported temperature to differ from the theoretical temperature of boiling water at her altitude. Under C-OAR-SE, the (accidental) empirical test that uncovered a serious validity problem would not even be an option, leaving our cheesemaker "expert judge" to wonder why she continued to fail.

We prefer a comprehensive approach to research, such as that described in Bagozzi's (1984; Bagozzi and Phillips, 1982) Holistic Construal. Such an approach recognizes that theory-building research involves three distinct domains – the conceptual domain, the empirical domain, and the domain of correspondence rules,

which specify how the conceptual links to the empirical. Careful consideration of correspondence rules alerts the researcher to potential confounds related to unreliability and invalidity. Not one of these three domains can be safely ignored or dismissed with simplistic assertions without grave danger to a study's conclusions.

C-OAR-SE reinforces an existing weakness of marketing measurement

As we have already alluded to, we view the expression of C-OAR-SE in MMR as a reinforcement of the dogmatic approach to measurement which has already harmed progress in marketing and social science research for so long. This reinforcement is ironic, since we believe Rossiter shares our goals of advancing marketing measurement. Yet, just as what Rossiter perceives as the psychometric paradigm is not appropriate to all measurement situations, neither is C-OAR-SE. And just as Rossiter is correct to point out that blanket acceptance of a single model has harmed us in the past, we are also correct to point out that replacing one dogma with another has uncomfortable connotations of the misguided removal of babies along with bathwater. In fact, as we read Churchill (1979), it is clear that he did not insist that all marketing measurement follow the proposed paradigm. Instead, Churchill asserted only that the paradigm had worked well on several occasions. For our part, we absolve Churchill, and place the blame squarely on the editors, reviewers, and researchers who turned Churchill's recommendations into a caricature. Yet this is by no means unique to measurement theory. For example, thousands of researchers have cited Armstrong and Overton (1977) to support claims of sample representativeness, without actually applying their procedures for estimating non-response bias. Even more broadly, applied researchers took pieces of the distinct and incompatible approaches to hypothesis testing developed by Ronald Fisher, on the one hand, and Jerzy Neyman and Egon Pearson, on the other, and fused them together into one hybrid approach to hypothesis testing, which is applied ritualistically (Gigerenzer, 2004; Gigerenzer *et al.*, 1989). But even in light of this clear tendency for applied researchers to turn recommendations into dogma, MMR explicitly demands absolute conformity to its version of orthodoxy, turning a tendency into a mandate.

A misdirected attack on a thirty-year-old classic

We suspect that Rossiter (along with the majority of experts) would agree with us in the opinion that Churchill's (1979) paper was a monumental achievement at the time and in its context. However, it is now more than 30 years old. Few fields stand where they did in 1979 – the era of the Cold War, and before the launch of the personal computer and the World Wide Web.

However, despite the progress made in many areas, a survey of contemporary measurement practices in marketing research will reveal a high degree of superficial similarity with Churchill's three-decade-old recommendations. The one substantial deviation might be the recent growth of interest in formative measures, an approach largely consistent with Rossiter's recommendations. Indeed, Rozeboom's (1960, p. 417) observation about psychologists seems appropriate enough for marketing researchers:

> [. . .] the perceptual defences of psychologists are particularly efficient when dealing with matters of methodology, and so the statistical folkways of a more primitive past continue to dominate the local scene.

And yet, Rossiter's recommended procedure sounds more like a return to practices denounced in Jacoby's (1978) critical review, as cited in Churchill (1979). C-OAR-SE promotes a return to a time when "most of our measures are only measures because someone says that they are, not because they have been shown to satisfy standard measurement criteria (validity, reliability, and sensitivity)" (Jacoby, 1978, p. 91).

While Churchill would doubtless be the first to admit that the state of the art in measurement has progressed beyond his 1979 summary, it seems unwarranted for MMR to contain such a scathing critique of his seminal paper. Certainly, researchers should make certain that their own work stands up to similarly close scrutiny. Thus, while we see no value in addressing MMR's critique of Churchill (1979) point-by-point, it seems only fair to identify a number of specific areas that appear somewhat inconsistent or even incorrect.

First and perhaps foremost, despite the claims in MMR to the contrary, it is hard to escape the feeling that Rossiter has set up a number of straw men for criticism. For example, it seems inconsistent for MMR to strongly criticize the reliance of researchers on the guidelines of Churchill (1979), and then to criticize researchers for not adhering to those same recommendations for minimum recommended alpha scores. Coefficient alpha itself is a clear example of a straw man, given the large amount of research that has already advocated caution in the use of alpha (e.g. Green and Yang, 2008).

There also appears to be some confusion over the concept of measurement error. We would argue that Churchill's (1979, p. 7) "systematic error" concept is simply what is called "specific error" in the factor analysis literature (i.e. the reliable score component specific to a particular indicator). Thus, it is hard to see how it differs from the concept of "measure-induced distortion" introduced in MMR. Further, while we applaud the attention given to test-retest reliability in MMR, the concept is not dealt with in a logical way. In particular, it is easy to dispute the claim on p. 15 of MMR that test-retest correlations must be demonstrable over a short interval. It seems that if a construct is malleable, or fleeting, the test-retest correlation will be uninformative. Further, the coverage of test-retest validity seems to be inconsistent with the emphasis on expert judgment evident in MMR; if experts agree the measure is valid, why is test-retest validity relevant in the C-OAR-SE model? And if it is relevant, why are other empirical tests not relevant?

Finally, MMR appears to assert an idiosyncratic definition of "psychometrics," identifying it with a particular type of model – apparently, the factor model – instead of with a discipline. However, although the factor model and associated techniques occupy an important place in psychometrics, they do not identify the discipline.

What binds psychometrics is the study of how theoretical attributes may be connected to empirical observations in psychology; this may in fact be called the central question of psychometrics. The factor model is associated with one particular answer to this question – namely, that items or test scores may be viewed as measures of a continuous latent variable, contaminated by measurement error. However, other models and answers to the question exist, as even a cursory glance at the psychometric literature shows. Most pertinent to the present discussion is the implication that the ideas in MMR cannot logically be placed in opposition to psychometrics, suggested in MMR. Instead, MMR poses a different answer to the central question of psychometrics,

and therefore sits squarely within the field. In other words, MMR presents a psychometric theory, whether its author likes it or not.

What can we do about it?

We have tried to make clear that our views overlap with those expressed in MMR in a number of ways regarding the extant problems in marketing and social science measurement. However, our manifesto for solving these problems is considerably different. Specifically, we believe the responsibility for ensuring that rigorous measure development in marketing and social science does not rest primarily with that minority of scholars who – like Rossiter and ourselves – take a specific interest in measurement as a field of study. Rather, we believe it is incumbent on individual researchers, editors, and reviewers to ensure the propagation of appropriate usage of measurement models and tools, rather than allowing perpetuation of narrow dogma.

We have cited three specific areas of concern regarding measurement practice in marketing: unthinking imitation of popular measurement approaches, overreliance on statistics at the expense of careful conceptualization, and failure to look beyond a narrow range of measurement alternatives. Taken together, they point to measurement practice on autopilot. We understand the pressures facing publication-minded researchers. Some measurement approaches increase scrutiny by reviewers and editors, and some do not. Some approaches are well-supported by computational tools and by instructional texts, and some are not. Some issues must be given their share of researchers' scarce time and attention, while other issues can be sacrificed at little cost. In this way, marketing as a discipline gets the measurement practice that it rewards.

At the same time, we believe in marketing researchers as scientists. These marketing scientists skeptically question other aspects of their research methodology, alert for confounds, and they similarly have the ability to critically examine their measurement practice. If reviewers and editors routinely and consistently asked authors to justify their measurement models, researchers would quickly get the message. We (like Churchill, and indeed surely Rossiter) think that these questions start with definitions of constructs. Researchers must conceptualize their constructs at a far more elemental level, and measure them accordingly, rather than attempting to create single constructs for what are clearly multiple and distinct phenomena. Certainly, if this is done, theories (and models) will be more complex. But complex theories with measurement at a more elemental level do more to advance knowledge than simple theories that use overly-complex constructs. Parsimony is only a useful goal insofar as it does not obscure useful representation of reality.

Questions about researchers' choice of measurement model must also encompass not only the rationale for a particular choice, but what kind of evidence would invalidate the researchers' choice, and what consequences might follow if the researchers are wrong. The position in MMR is that the dire consequences exist, but that the evidence does not. We argue that impossibility of invalidation is a sign of faith, not scientific method. In this sense, reviewers (and researchers) would do well to learn from the methodological literature first-hand, rather than relying on second-hand dogma and yielding to myth and misunderstanding.

Further, reviewers must make efforts to fully understand the appropriate use of measurement models. Far too often are inaccuracies perpetuated in even our best

journals. It seems often that reviewers are willing to overlook such methodological problems if the theoretical story of a paper is compelling. However, how can strong empirical conclusions be possible without good measurement of the variables in that theory? We need the same rigorous questions across all three domains of research – the conceptual, then empirical, and the correspondence rules which link the two.

We also entreat journal editors to encourage research whose aim is to broaden and deepen our understanding of measurement options. Measurement dogma suffocates research on measurement, insisting that all answers to measurement questions are believed to lie within a canon of respected publications. Journal editors are in position either to deter innovation by dismissing whole areas of inquiry, or to spur inquiry by highlighting issues of importance. We do not believe that reflective measurement, formative measurement, or Rossiter's C-OAR-SE, are the only viable perspectives on measurement in marketing. We hope many marketing scholars would welcome a more open conversation.

Finally, those who look to contribute to measurement theory also bear some responsibility. Most importantly, we must not confuse weak applications of a measurement model with inherent weaknesses of the measurement model. In exploring the vast territory of plausible measurement models, researchers will no doubt make mistakes. Nevertheless, to improve the quality of measurement in marketing and social science over the long term, our discipline needs to encourage a greater set of measurement models to be considered by researchers, not demand dogmatic adherence to any single model.

References

Apel, H. and Wold, H. (1982), "Soft modeling with latent variables in two or more dimensions", in Jöreskog, K. and Wold, H. (Eds), *Systems under Indirect Observation*, Vol. 2, North-Holland, Amsterdam, pp. 209-48.

Asparouhov, T. and Muthén, B. (2010), "Bayesian analysis of latent variable models using Mplus", unpublished manuscript, available at: www.statmodel.com/download/BSEMv4.pdf (accessed December 28, 2010).

Bagozzi, R.P. (1984), "A prospectus for theory construction in marketing", *Journal of Marketing*, Vol. 48, Winter, pp. 11-29.

Bagozzi, R.P. and Phillips, L.W. (1982), "Representing and testing organizational theories: a holistic construal", *Administrative Science Quarterly*, Vol. 27, pp. 459-89.

Bartholomew, D.J., Deary, I.J. and Lawn, M. (2009), "A new lease of life for Thomson's bonds model for intelligence", *Psychological Review*, Vol. 116, pp. 567-79.

Borsboom, D. (2006), "The attack of the psychometricians", *Psychometrika*, Vol. 71, pp. 425-40.

Borsboom, D. (2008), "Psychometric perspectives on diagnostic systems", *Journal of Clinical Psychology*, Vol. 64, pp. 1089-108.

Chaline, E. (2009), *History's Worst Inventions*, Fall River Press, New York, NY.

Churchill, G.W. (1979), "A paradigm for development of better measures of marketing constructs", *Journal of Marketing Research*, Vol. 16, February, pp. 64-73.

Cliff, N. (1983), "Some cautions concerning the application of causal modeling methods", *Multivariate Behavioral Research*, Vol. 18, January, pp. 115-26.

Cohen, J. (1982), "Set correlation as a general multivariate data-analytic method", *Multivariate Behavioral Research*, Vol. 17, July, pp. 301-41.

Cramer, A.O.J., Waldorp, L.J., van der Maas, H. and Borsboom, D. (2010), "Comorbidity: a network perspective", *Behavioral and Brain Sciences*, Vol. 33, pp. 137-93.

Edwards, J.R. (2001), "Multidimensional constructs in organizational behavior research: an integrative analytic framework", *Organizational Research Methods*, Vol. 4, pp. 144-92.

Franke, G.R., Preacher, K.J. and Rigdon, E.E. (2008), "Proportional structural effects of formative indicators", *Journal of Business Research*, Vol. 61, December, pp. 1229-37.

Gigerenzer, G. (2004), "Mindless statistics", *The Journal of Socio-Economics*, Vol. 33, pp. 587-606.

Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J. and Krüger, L. (1989), *The Empire of Chance: How Probability Changed Science and Everyday Life*, Cambridge University Press, Cambridge.

Green, S.B. and Yang, Y. (2008), "Commentary on coefficient alpha: a cautionary tale", *Psychometrika*, Vol. 74, pp. 121-35.

Howell, R.D., Breivik, E. and Wilcox, J.B. (2007), "Reconsidering formative measurement", *Psychological Methods*, Vol. 12, pp. 205-18.

Jacoby, J. (1978), "Consumer research: a state of the art review", *Journal of Marketing*, Vol. 24, April, pp. 87-96.

Lee, N. and Hooley, G.J. (2005), "The evolution of 'classical mythology' within marketing measure development", *European Journal of Marketing*, Vol. 39, pp. 365-85.

McDonald, R.P. (2003), "Behavior domains in theory and in practice", *Alberta Journal of Educational Research*, Vol. 49, pp. 212-30.

McGrath, R.E. (2005), "Conceptual complexity and construct validity", *Journal of Personality Assessment*, Vol. 85, pp. 112-24.

McGrath, R.E. (2009), "On prototypes and paradigm shifts", *Measurement*, Vol. 7 No. 1, pp. 27-9.

Mershon, B. and Gorsuch, R. (1988), "Number of factors in the personality sphere: does increase in factors increase predictability if real life criteria?", *Journal of Personality and Social Psychology*, Vol. 55, pp. 675-80.

Mulaik, S.A. and Quartetti, D.A. (1997), "First order or higher order general factor?", *Structural Equation Modeling*, Vol. 4 No. 3, pp. 193-211.

Muthén, L.K. and Muthén, B.O. (2010), *Mplus User's Guide*, 6th ed., Muthén and Muthén, Los Angeles, CA.

Paunonen, S.V. (1998), "Hierarchical organization of personality and prediction of behavior", *Journal of Personality and Social Psychology*, Vol. 74, pp. 538-56.

Pearl, J. (2010), "The causal foundations of structural equation modeling", Technical Report, Computer Science Department, University of California, Los Angeles, CA.

Reise, S.P., Widaman, K.F. and Pugh, R.H. (1993), "Confirmatory factor analysis and item response theory: two approaches for exploring measurement invariance", *Psychological Bulletin*, Vol. 114 No. 3, pp. 552-66.

Rigdon, E.E. (1994), "Demonstrating the effects of unmodeled random measurement error", *Structural Equation Modeling*, Vol. 1 No. 3, pp. 375-80.

Rindskopf, D. and Rose, T. (1988), "Some theory and applications of confirmatory second-order factor analysis", *Multivariate Behavioral Research*, Vol. 23, pp. 51-67.

Rossiter, J.R. (2002), "The C-O.A.R-S.E.procedure for scale development in marketing", *International Journal of Research in Marketing*, Vol. 19 No. 4, pp. 305-35.

Rossiter, J.R. (2008), "Content validity of measures of abstract constructs in management and organizational research", *British Journal of Management*, Vol. 19 No. 4, pp. 380-8.

Rossiter, J.R. (2011), *Measurement for the Social Sciences: The C-OAR-SE Method and Why It Must Replace Psychometrics*, Springer, New York, NY.

Rozeboom, W.W. (1960), "The fallacy of the null-hypothesis significance test", *Psychological Bulletin*, Vol. 57 No. 5, pp. 416-28.

Schneider, R.J., Hough, L.M. and Dunnette, M.D. (1996), "Broadsided by broad traits: how to sink science in five dimensions or less", *Journal of Organizational Behavior*, Vol. 17, pp. 639-55.

Van der Maas, H.L.J., Dolan, C.V., Grasman, R.P.P.P., Wicherts, J.M., Huizenga, H.M. and Raijmakers, M.E.J. (2006), "A dynamical model of general intelligence: the positive manifold of intelligence by mutualism", *Psychological Review*, Vol. 113, pp. 842-61.

Corresponding author
Nick Lee can be contacted at: n.j.lee@aston.ac.uk