Welcome to the desert of the real: reality, realism, measurement, and C-OAR-SE

This item was submitted to Loughborough University's Institutional Repository by the/an author.


Additional Information:

- This paper was published in the journal European Journal of Marketing and the definitive published version is available at https://doi.org/10.1108/EJM-10-2016-0549.

Metadata Record: https://dspace.lboro.ac.uk/2134/22933

Version: Accepted for publication

Publisher: © Emerald

Rights: This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: https://creativecommons.org/licenses/by-nc-nd/4.0/

Please cite the published version.
WELCOME TO THE DESERT OF THE REAL: REALITY, REALISM, MEASUREMENT, AND C-OAR-SE

<table>
<thead>
<tr>
<th>Journal:</th>
<th>European Journal of Marketing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID:</td>
<td>EJM-10-2016-0549</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Original Article</td>
</tr>
<tr>
<td>Keywords:</td>
<td>Psychometrics, Realism, Measurement, Validity</td>
</tr>
</tbody>
</table>
WELCOME TO THE DESERT OF THE REAL: REALITY, REALISM, MEASUREMENT, AND C-OAR-SE

Nick Lee
Professor of Marketing
Warwick Business School
Warwick University

John Cadogan
Professor of Marketing
School of Business and Economics
Loughborough University

And

Professor of Marketing
LUT School of Business and Management
Lappeenranta University of Technology
Finland
Abstract:

Purpose – This article provides a balanced commentary on Rossiter’s paper ‘How to use C-OAR-SE to design optimal standard measures’ in this issue of the ‘European Journal of Marketing’. It also relates the comments in general to Rossiter’s other C-OAR-SE work, and throws light on a number of key measurement issues that seem under-appreciated at present in marketing and business research.

Design/methodology approach – We use conceptual argument based on measurement theory and philosophy of science.

Findings – We find that Rossiter’s work makes a number of important points that are necessary in the current stage of development of marketing and social science. However, we also find that many of these points are also well made by fundamental measurement theories. When measurement theory is correctly interpreted, the idea of multiple measures of the same thing is not problematic. However, we show that existing social science measurement practice rarely takes account of the important issues at play here.

Practical implications – We show that marketing, management, and social science researchers need to get better in terms of their appreciation of measurement theory, and in their practices of measurement.

Originality/value – We identify a number of areas where marketing and social science measurement can be improved, taking account of the important aspects of C-OAR-SE, and incorporating them in good practice, without needlessly avoiding existing good practices.

Keywords Psychometrics; Realism; Measurement; Validity
Introduction

For almost two decades, we have enjoyed various jousts with John Rossiter over measurement issues, and particularly his C-OAR-SE theory / method. We read early drafts of the paper that was eventually published in the *International Journal of Research in Marketing* (Rossiter, 2002), we delivered opposing presentations on measurement in special conference sessions (e.g. the *European Marketing Academy 2005*) and have also written commentaries on C-OAR-SE (e.g. Rigdon et al., 2011, and this present comment). We’ve relished informal discussions during conference sessions, and over drinks later on, as well as debates over email on occasion. Through it all has been John’s constant challenge to improve our measurement practices in marketing and wider social sciences. Whether or not C-OAR-SE has been significantly impactful on the practice of marketing measurement (and Rossiter in this issue implies not), it is hard to deny the impact it has had on driving those who, like us, take a specific interest in measurement, to sharpen their own theories and methods, even if the purpose of which was sometimes to more strongly rebut Rossiter’s arguments.

Our goal in this present comment is to place Rossiter’s (2016) article in the *European Journal of Marketing* – which we will refer to as JR2016 – in a broader context of what we term metaphysical measurement theory\(^1\). In doing so, we will clarify the areas where the principles outlined in C-OAR-SE continue to have the potential to make a positive impact on scholarly practice. One of our primary goals is to explore

---

\(^1\) We use this term to refer to a measurement theory that specifically focuses on questions concerning the fundamental structure of the world. Such issues include ontological questions about the identity, and reality, of the things we try to measure (especially unobservables), as well as causality. We would consider ‘empirical measurement theory’ by contrast to primarily be concerned with the modeling of empirical data. While both are important, it strikes us that most existing social science measurement theory concerns itself much more with the latter than the former.
the areas where JR2016, and other C-OAR-SE theoretical work, identifies important symptomatic problems in our current practice, but at the same time does not diagnose the fundamental underlying conditions responsible for those symptoms. In such cases, C-OAR-SE on occasion prescribes remedies that may only go part way towards curing the problem. In particular, as we shall show here, many aspects of C-OAR-SE present what really should be good practice in existing measurement theory, in new and complex ways, which may be behind Rossiter’s lament that C-OAR-SE is seen as a ‘seemingly too difficult measure-design method’ (JR2016 PPno). We suggest some simplifications and clarifications that may enhance the take-up of the strong fundamental principles Rossiter has long advocated in C-OAR-SE.

Our mainly metaphysical focus means that we leave many of the more specific empirical and analytic aspects of JR2016 to our fellow commenters, and focus our attention on a single theme of JR2016, namely that there are too many “different, even slightly different, measures of the same construct” and, worse, that there is no guarantee that the use of different measures will not “produce different findings” (pp XX). We strongly agree with JR2016 that the multiple measures issue is in current practice a serious problem. However, our view is that this is a symptom of a much more serious underlying problem in social measurement practice, that of the general lack of convincing definitions of what the unobservable attributes (which are represented by constructs in social theories) being measured actually are, and whether they exist or not. This problem is compounded by weak theories about how these attributes are connected to their measuring devices. Below, we place a number of JR2016’s recommendations in context with these key foundations.
Multiple Constructs, Multiple Measures … Multiple Realities?

Rossiter has long been concerned about the seemingly ever-growing variety of measures created and utilized across different studies, which purport to ‘measure’ the same construct (e.g. Rossiter, 2002, 2011). In JR2016, he uses the well-known construct of ‘market orientation’, among others, to illustrate this point, although we could point out a seemingly endless variety of others. We cannot do anything other than agree with Rossiter here. For too long, the social sciences have all contributed to an ever-growing menagerie of ‘constructs’, each then often ‘measured’ using multiple different operationalizations. This is by no means limited to management and marketing, and even a cursory search of literature in fields such as psychology, education, sociology, and suchlike will show multitudes of different constructs, many offering slight variations on an already-existing theme, or even giving the same name to a different definition, let alone the manifold different measures of ostensibly the same construct.

Time and time again, we see the reasoning behind such a situation based on something called ‘classical test theory’, ‘latent construct / variable theory’, the ‘domain sampling model’, or ‘psychometric theory’. However, there seem to be a number of quite significant inconsistencies between a) what various scholars mean by these terms, b) what they actually do in practice, and c) what the relevant underlying measurement theory really is. Therefore at this point it is necessary to clarify some terms, to avoid confusion. In particular, the word attribute in C-OAR-SE is used to refer to only one part of a construct definition, and oftentimes, multiple attributes (which Rossiter terms ‘subattributes’) are combined in one construct definition. This seems to contradict most ideas of measurement (see Markus and Borsboom, 2013 for
an excellent overview), and instead implies that C-OAR-SE is a method to develop abstract composites, rather than measure actual features of the real world.

If indeed C-OAR-SE is intended to be a composite construction theory, it is of little use to think of it as a measurement theory (Edwards, 2011; Markus and Borsboom, 2013; Cadogan, Lee, and Chamberlain, 2013), and to subject it to scrutiny as such. However, we are relatively certain that Rossiter views C-OAR-SE as a measurement theory, given the copious references to measurement throughout C-OAR-SE papers. If so, we must be clear about what measurement itself actually is. We are not able to do more here than briefly sketch the issues, and hence for simplicity’s sake we follow Markus and Borsboom (2013) in using the term causal theory of measurement, which seems to be a “widely accepted set of assumptions among researchers” (Markus and Borsboom, 2013, pp. 84) regarding their activities when engaged in measurement practice. A causal theory of measurement (CTM) takes the constructs in measurement models to “refer to common causes (equivalently, attributes; Rozeboom, 1966) that underlie a set of item responses, so that people respond to items differently because they have a different construct score” (Markus and Borsboom, 2013, pp. 81-82).

As such, constructs are simply components of a scientific theory, which themselves have no existence, because a theory itself is simply an abstract representation. However, in a CTM, we must remember that constructs are designed to be referent devices which stand in for actual attributes of the real world we are studying. In other words, one does not measure a construct, one measures an attribute. This distinction shall prove critical to the following discussion, and is tentatively expressed in Figure 1a, which connects measurement theory with actual empirical practice. Importantly
Figure 1a avoids conflation of purely theoretical and purely empirical/mathematical terms (constructs and empirical scores) with actual features of the real world (attributes and measures). Such problematic intermingling is encoded in typical measurement model visualizations (see for example Figure 1b), and seems likely to have led to a number of unfortunate consequences for actual social measurement thought and practice.

FIGURE 1a and 1b.

The idea of a CTM is likely to seem so self-evident as to be almost trivial to many readers – after all, without ‘something’ to measure, how can we measure? However, encoded in the seemingly simple idea are a number of very powerful assumptions (as expressed in Figure 1a), which closer analysis of actual social science practice (more aligned with Figure 1b) will show to lack convincing support, leading to the symptoms Rossiter, in JR2016 and elsewhere, identifies (Rigdon 2016 also touches on similar issues, albeit for a slightly different purpose). Unfortunately, at times C-OAR-SE also falls into similar conceptual traps, even as it identifies important problems in current practice. We cannot address them all here, but the most critical issue concerns the nature of the attribute one is trying to measure. More specifically, are attributes real? A CTM seems to require the assumption that attributes being measured are real, otherwise, how could they exist, how could they have causal impacts on other attributes, and indeed how could they therefore be measured? Social scientists face the extremely difficult problem that most of the attributes they are interested in are unobservable (i.e. not detectable by the unaided human senses), and are therefore not subject to direct confirmation of their existence. However, social science is hardly
alone in this, and even modern physics faces similar hurdles, such as (to use the terms of the present discussion for simplicity’s sake) measuring the attributes of subatomic particles (e.g. electron spin) and suchlike.

A common response to this problem is to appeal to a realist philosophy, where unobserved attributes in scientific theories are considered to actually exist, as part of a mind-independent world that is investigated by the sciences (Chakravartty, 2015, see for example Borsboom, Mellenbergh, and Van Heerden, 2003), which would seem superficially to settle the issue. However, this sort of ‘blanket realist’ approach does not itself survive much scrutiny, and leaves a number of questions without satisfactory answers. One fundamental question concerns how we can reconcile a scientific realism with the fact that numerous unobservable attributes and entities have been debunked as science has progressed, which is termed the pessimistic induction objection (Chakravartty, 2015). If for example the phlogistication of air is no longer viably real, why are we confident that the market orientation of a firm, or even individual differences such as personality traits (or for that matter, unobservable physical attributes such as electron spin), are real? In other words, given that the reality of all past unobservable attributes that were posited in scientific theories has been disproved, by what logic can we consider the attributes referred to by today’s scientific theories to be real?

Most importantly in the present context however, a blanket realism seems to admit any and all of the attributes we propose to the pantheon of reality. This inevitably leads us to anarchy, which is essentially the situation described by JR2016. However, where JR2016 appears to problematize the proliferation of different measures, we
would instead argue that the real issue is a seemingly unstoppable propagation of
different constructs, many of which have the same name, even though they seem to
refer to different attributes, which themselves are only weakly defined. Market
orientation, for example, had multiple construct definitions almost as soon as it
appeared in the literature (e.g. Kohli and Jaworski, 1990; Narver and Slater, 1990),
and there have been a seemingly endless variety of new definitions in the last quarter-
century. So, what is ‘market orientation’? If it is real, how can there be multiple
different definitions of it? By basic logic, either there must be a) one thing, which we
call market orientation b) multiple different things, which should therefore be given
different names, or c) no thing. What there cannot be is multiple things that are the
same thing. The multiple measures JR2016 criticizes are a symptom of this absolutely
crucial problem, common to social sciences in general. A science cannot progress – or
even begin to call itself a science – if it cannot agree on what its fundamental
attributes are, let alone whether they actually exist.

One Reality, Multiple Measures, No (Necessary) Problem

If convincing descriptions of fundamental attributes can be provided, the multiple
measures issue pointed out by JR2016 ceases to be a significant problem. Take the
attribute of ‘temperature’ for example². Of course, temperature has the advantage of
being directly observable by our unaided senses, but the principle holds to all real

---
² We recognize that we are at risk of oversimplifying the physical complexities of even this basic
example. However, to clarify, ‘heat’ is the total energy of molecular motion of a given body, whereas
‘temperature’ is the average energy of molecular motion in that body. Thus, heat depends partly on the
size of the body, while temperature does not. Therefore, heat can be understood as energy, while
temperature can be understood as the average energy content of a body. Temperature has a specific and
well-defined theoretical relationship to heat, but is not the same thing, since two bodies of the same
substance but different size can have the same temperature but different amounts of heat. We therefore
refer to ‘temperature’ as an attribute of an object, which can be measured. Importantly, the temperature
attribute is an existent property of the real world, separate from the scale used to measure it (e.g.
Celsius, Kelvin). Thus, we abstract away from the less relevant (for our present purposes), although
interesting, controversies over the definition of temperature, and the creation of viable temperature
scales and points of reference (e.g. see Chang, 2004).
attributes, observable or not. Consider that one can indeed have multiple valid
temperature measures, such as a simple mercury thermometer, a gas thermometer, an
infrared thermometer, or even the complex spectral radiance approaches used to
measure the cosmic background radiation. One assumes that Rossiter does not
consider this to be problematic in itself. As such, our view is that the basic principle
of multiple measures of the same thing is not itself problematic, providing that we can
convincingly argue that we are measuring a real attribute.

Unfortunately, we are a very long way from having a set of convincing descriptions of
the fundamental attributes that social science in general studies, let alone marketing.
In this case, multiple measures are a major practical problem, leading to serious
problems with research comparability, analogous to if a mercury and IR thermometer
each actually measured different attributes, neither of which we were sure really
existed, and yet we referred to both as ‘temperature’ measures. In fact, the situation in
social science may be even worse, because it is surprisingly difficult to even conceive
of how we could ever come up with strong and convincing descriptions of even the
most fundamental human attributes, beyond the physical (cf. Nagel 2002). For
example, do mental experiences such as love, happiness, pleasure, and other things
we might call ‘feelings’, exist separately from their biological and physical
manifestations? If we cannot provide supportable arguments for even this, where does
that leave us in terms of the social attributes referred to by constructs such as power,
or charisma? And what then becomes of market orientation, service quality, or brand
equity? Under what criteria do we justify these as ‘real’?
Although they are serious, it is outside the scope of this present paper to do more than raise these issues, and note that we have seen very little discussion of anything beyond a blanket realism in recent social science measurement theory (with the notable exception of a brief discussion in Markus and Borsboom, 2013, pp 308-312).

So, for the purposes of continuing the discussion, let us make the (somewhat heroic) assumption that we can indeed provide some convincing descriptions of real marketing-relevant attributes, such as, say, attitude towards an advertisement. If we are to maintain a convincing measurement theory, we must also then set out a strong theory as to how the attribute causes changes in the measuring device. Again, temperature is a useful example. There is a clear theory which shows how the attribute of temperature of a substance causes some reaction of the measuring device (usually based around the kinetic theory, although statistical physics / quantum mechanical explanations are also available). If this is the case, again, multiple measures *per se* are not a problem, because it is possible to validate each measuring device separately, and convergent validation becomes a mere (and somewhat meaningless) support to this. It would after all be rather disturbing if an IR and mercury thermometer turned out significantly different results.

Again though, in social sciences like marketing we are a long, long, way from this. Unfortunately, the content validity idea, even wrapped up in the ‘construct-to-measure’ (CtM) terminology JR2016 uses, does not suffice alone. Agreement amongst experts that, for example, some questionnaire item ‘should’ measure some unobservable property is definitely not equivalent to the theoretical bases for temperature measurement (observable), or electron spin (unobservable) for example.

We recognize that we are setting what some may consider an impossibly high bar for
marketing measurement, but our intention is more to show how far we have to go, rather than the fact that the destination is unobtainable. Indeed, even something as fundamental as temperature measurement was controversial only around 150 years ago (Chang, 2004). As such, we certainly agree that comprehensive approaches to delineating the theoretical (which is by definition causal) link from the unobservable property to the measure (Link a in Figure 1a) are essential to basic good practice in measure development.

We also agree with Rossiter that universal appeal to multi-item convergent correlations is not a substitute for a strong causal theory of how an unobservable attribute is measured. This is in fact implied in Figure 1a. Convergent validation essentially concerns comparison of empirical scores, which Figure 1a shows have no direct connection to the unobservable attributes one is trying to measure. If there are problems with the other links in the Figure (which is very likely given our current state as described above), then comparing empirical scores will be meaningless evidence of whether two measures are equivalently valid. Of course, one would expect multiple measures of the same thing to be correlated (consider, again, the example of temperature), but alone this does not provide convincing evidence of what has been measured, or even if anything has been measured at all. Indeed, the value of any empirical evidence rests on the strength of the underlying theory that provides the hypotheses that are being tested. Where we disagree with Rossiter is that we believe any coherent theory of measurement must as a key feature provide strong and falsifiable hypotheses about empirical data. Without even the possibility of empirical testing, it is hard to consider any theory to be scientific.
That said, without a strong theory, any empirical results are essentially meaningless in terms of supporting the existence of unobservable properties. The problem we face in social science, which is the root cause of the symptoms JR2016 identifies, is that we have very little theory of significant strength in social science measurement regarding how measures should behave. Appealing to a conditional independence hypothesis (i.e. that correlations between multiple items vanish when conditioned on a common cause) is about the strongest we can do right now, with more than one so-called ‘measurement’ theory unable to even do that (see Markus and Borsboom, 2013).

However, a strong theory should also rule out alternative explanations for the observed pattern of empirical data. Current social science measurement theories do not do a good job of this, since they are reliant solely on observable correlation in the most part, and do little to explain how the purported unobservable property has a causal impact on the measurement device(s). As such, multiple alternative explanations for the observed data are available. Again, to use temperature as the example, clear mechanisms are theorized about why different temperature measures change in response to changes in the temperature of a substance. This then explains why the individual measures would likely be correlated. The correlation itself is not taken as evidence temperature exists. Of course, temperature is observable, but unobservability itself is not an insurmountable barrier to strong measurement theories.

If it were, most of modern physics, for example, would be untenable.

Summary and Conclusions

We understand the intention behind C-OAR-SE, and we fully support it. Indeed, we agree that measurement theory in marketing and other social sciences is in a worse than parlous state, and without a coherent measurement theory, there can be no
Claiming that social science is spared this task because it deals with abstract and unobservable attributes is no defence, because almost all modern sciences must face in one way or another such issues. Social science is not unique in dealing with unobservables, even if it does face numerous complex and interrelated challenges. Further, we agree with Rossiter that in current practice, the seemingly unstoppable propagation of multiple measures with the same name is a serious problem. We also have strong support for Rossiter’s efforts, both here in JR2016 and elsewhere, to increase recognition that the theoretical description of a measured attribute, and the theoretical validation of the measure, are important considerations which are seriously under-appreciated at present, and we take many of his C-OAR-SE principles in this spirit.

So, why, as Rossiter himself bemoans in JR2016, has C-OAR-SE remained stubbornly on the margins of current practice? Our view is that this is because in some ways we think C-OAR-SE goes too far in its suggested remedies for the symptoms it identifies, analogous to amputating a limb when a bone is broken. Conversely, we also think C-OAR-SE does not go far enough in diagnosing the real underlying issues which are driving poor practice, analogous to diagnosing a muscle sprain when the bone is broken.

In our view, the underlying problem with social science measurement is the at best weak, and at worst fundamentally unsound, state of current theorizing about the unobservable attributes we are attempting to measure. Following from this is the

---

3 The most famous expression of this view is by Lord Kelvin, who stated “when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science” (Thomson, 1889, pp. 73).
preponderance of implausible and even nonsensical measurement operationalizations.

Sole reliance on convergent validation and correlation, as JR2016 correctly points out, is certainly problematic practice in light of these underlying problems, and provides little convincing evidence of measurement validity, or even if there is something to measure.

However, we cannot agree that a blanket rejection of all empirical testing of measures, as Rossiter in JR2016 and elsewhere advocates, is a plausible solution. Even if it were, the state of our current knowledge about the social and human behavioral world is nowhere near detailed enough to give us the ability to create strong enough theories connecting most unobservable attributes of relevance to social science to observable measures, in the way done with unobservable physical attributes such as electron spin, or even observable attributes such as temperature. We must therefore do the best we can, and continue to strive for improvement. Our strong theories must have the potential to make clear and falsifiable empirical predictions, otherwise, what are they theories of? However, as well as this, they must do better at ruling out alternative explanations. Doing so is the job currently of a strong interplay between theoretical mechanisms and patterns of empirical observations. Ignoring either one leaves us in a highly dangerous state. Rossiter is correct in JR2016 to criticize unqualified reliance on observed correlations as evidence of valid measurement. However, replacing this with what seems to be an almost totalitarian reliance on expert judgement, as CtM validity seems to advocate, seems to us to be equally unreasonable.

To conclude, while C-OAR-SE may not have assumed the dominant place in the methodological toolkits of marketing and social scientists that Rossiter hoped, it is
undeniable that it has had a significant influence on the development of measurement theory itself. This has been recognized through multiple awards, including the 2012 Jan-Benedict E. M. Steenkamp Award for Long-Term Impact, and the 2011 European Journal of Marketing Best Paper Award. However, C-OAR-SE has also played a very significant role in bringing an expanded understanding of measure development to marketing researchers, and encouraged many to think outside what are often accepted as the ‘standard’ approaches to developing a marketing measure. While C-OAR-SE as a self-contained theory itself may never become as popularly-used by practicing researchers as Rossiter hopes, many of its key principles deserve to be taken seriously. We hope this present commentary encourages further development of these principles, with the end goal of a fully-realised and robust theory for social science research. We are a long way from that at present, but without authors such as Rossiter (and the other commenters in this issue), we would be far further away. And to be absolutely clear, without measurement, there can be no social science.
REFERENCES


FIGURE 1a

Note: The issue of error in these links is left unexplored at present. We assume that links B) and A) will have some level of error, but links C) and D) require more detailed explication in this regard. This is left for future consideration.