



Contents lists available at ScienceDirect

## Personality and Individual Differences

journal homepage: [www.elsevier.com/locate/paid](http://www.elsevier.com/locate/paid)

# Personality assessment, ‘construct validity’, and the significance of theory<sup>☆</sup>

Simon Boag

Department of Psychology, Macquarie University, Sydney, NSW 2109, Australia

## ARTICLE INFO

## Article history:

Received 12 September 2014  
 Received in revised form 16 December 2014  
 Accepted 22 December 2014  
 Available online xxx

## Keywords:

Latent variables  
 Likert scales  
 Measurement  
 Personality assessment  
 Personality traits  
 Psychometrics  
 Realism  
 Theoretical variables

## ABSTRACT

Personality assessment helps us to predict how people behave under various circumstances or how well a person might perform within certain roles. However, there are reasons to question the supposed ‘construct validity’ of tests designed to assess various personality attributes including dispositional traits. To demonstrate this, the paper first discusses a realist account of test validity where validity requires that both the attribute exist and that changes in the attribute are causally related to changes in test scores. The paper demonstrates that the validity for tests of dispositional traits is questionable given conceptual problems with traits existing as within-person attributes capable of causing changes in test scores. The widespread reliance on Likert-style response formats is then discussed in relation to the assumed quantitative structure of personality attributes. Based on a realist view of measurement, the uncritical adoption of a representational theory of measurement within personality research means that the validity of all personality tests claiming to ‘measure’ personality attributes is questionable. Suggestions for addressing test validity in personality assessment are then discussed in terms of paying greater critical attention to personality theory itself and adopting a realist theory of assessment and measurement.

© 2014 Elsevier Ltd. All rights reserved.

## 1. Introduction

Personality assessment currently helps predict how people behave under various circumstances or how well a person might perform within certain roles (e.g., Barrick, Mount, & Judge, 2001; Berry, Ones, & Sackett, 2007; Furnham & Fudge, 2008; Salgado, 2003). However, what personality tests are *actually* assessing is not entirely clear and whether personality variables such as dispositional traits are even sensibly formulated remains a matter of dispute. The specific issue here concerns the clarity of personality attributes such as dispositional traits as found in the Five Factor Theory (FFT) and Model (FFM) (Costa & McCrae, 1995, 2008, 2009). Long-standing arguments propose that dispositional traits are descriptive summaries reified into within-subject attributes and then erroneously used to explain individual behaviour (Boag, 2011a, 2011b; Cervone, 1999, 2004, 2005; Kroger & Wood, 1993). Consequently, whether personality assessment procedures assess

what they purport to assess is questionable because the understanding of the personality attribute itself is confused.

The aim of this paper is to address the issue of validity in personality assessment procedures and to demonstrate the significance of theory for developing valid personality tests. The paper first discusses Borsboom, Mellenbergh, and van Heerden’s (2004) realist account of test validity. This realist position is applicable to any personality tests claiming to assess actual personality attributes, including tests of dispositional traits, where traits are hypothesised to be attributes that are causally reflected in personality test scores (e.g., Costa & McCrae, 1995, 2008, 2009). While there are limitations with Borsboom et al.’s (2004) account with respect to assuming that attributes are quantitative (Michell, 2009b), a realist account of test validity provides two criteria for valid assessment: (i) the attribute in question exists, and; (ii) that differences in the attribute are causally reflected in differences in scores upon the assessment outcomes. As will be developed, the immediate problem with the assessment of dispositional traits is that neither (i) nor (ii) are satisfied: the reality of dispositional traits is questionable given logical problems in their conceptualisation. Specifically, dispositional trait accounts appear to reify person–situation relationships into attributes possessed by individuals and the subsequent problem of circularity precludes any causal influence of traits upon test performance (Boag, 2011a, 2011b). The paper then turns its attention to the broader issue of

<sup>☆</sup> An earlier version of the paper was presented at the 13th Australian Conference on Personality and Individual Differences (ACPID) hosted by Newcastle University, November 28–29, 2014. I would like to thank three anonymous reviewers for their valuable comments.

E-mail address: [simon.boag@mq.edu.au](mailto:simon.boag@mq.edu.au)  
 URL: <http://www.simonboag.com>

measurement in personality research and the lack of attention to the scientific task of quantification. The widespread reliance on self-report, Likert-style response formats is discussed in relation to the assumed quantitative structure of personality attributes. Based on arguments developed by [Michell \(1999, 2008, 2009b\)](#), the validity of all personality tests claiming to ‘measure’ personality attributes is questionable given the uncritical adoption of a representational theory of measurement within personality research. Suggestions for addressing test validity in personality assessment and measurement are then discussed in terms of paying greater critical attention to personality theory itself and adopting a realist theory of measurement.

## 2. What is meant by ‘construct validity’?

In their influential text on statistics and mental tests [Lord, Novick, and Birnbaum \(1968\)](#) claim that “[f]or scientific purposes, the most important characteristic of a test is its *construct validity*” (p. 278, their italics) and later [Clark and Watson \(1995\)](#) write that “[i]t has become axiomatic that (publishable) assessment instruments are supposed to be reliable and valid” (p. 309). Presumably all modern day personality test constructors would make some claim that their tests purport to assess some actual feature of persons. In this sense most researchers can be described as realists insofar as they accept that personality attributes exist, which provides a basis for the common view of ‘construct validity’—whether the test validly assesses what it is meant to assess ([Borsboom, Cramer, Kievit, Zand Scholten, & Franic, 2009](#); [Borsboom et al., 2004](#)). Indeed, whether acknowledged or not, realism underlies the scientific aim of discovering facts about the world and any subsequent psychometric practice ([Haig, 2014](#); [Hood, 2013](#); [Petocz & Newbery, 2010](#)). Unlike [Cronbach and Meehl’s \(1955\)](#) account of construct validity, however, where validity concerns how a ‘construct’ relates to theoretical and observational terms within a nomological network (rather than directly considering the nature of the attribute itself), a realist account of test validity places the burden upon the ontological status of the attribute of interest and its causal role in producing variations in test scores ([Borsboom et al., 2004, 2009](#); see also [Hood, 2009, 2013](#)). What is important here for test validity is the relationship of the attribute to the test score: validity involves a relation, and specifically a causal relation between an attribute *X* and the scores on some assessment procedure *Y*. As a causal theory both the attribute must exist independently of the assessment procedure and differences in the attribute must be responsible for differences (either directly or indirectly) in scores upon the assessment procedure (in a similar manner to how changes in actual temperature lead to changes in the rise and fall of mercury within a thermometer) ([Borsboom et al., 2004, p. 1066](#); cf. [Cramer et al., 2012a, p. 414](#)). None of this should be particularly controversial for most personality trait researchers since traits are taken to be real attributes of persons that are causally reflected in trait-assessment procedures (e.g., [Costa & McCrae, 1995, 2008, 2009](#)). Nevertheless, the situation is further complicated by the array of terms used (including ‘constructs’, ‘factors’, ‘latent variables’, ‘theoretical variables’, etc.), and while such terms are often treated as synonyms, it is not at all clear that they in fact are (see, for instance, [Maraun & Gabriel, 2013](#)).

This realist view of test validity provides two fairly obvious criteria for assessing the validity of any given test, the first of which (that the attribute exists) is necessary for the second to obtain:

If something does not exist, then one cannot measure it. If it exists but does not causally produce variations in the outcomes of the measurement procedure, then one is either measuring nothing at all or something different altogether. Thus a test is valid for measuring an attribute if and only if (a) the attribute

exists and (b) variations in the attribute causally produce variations in the outcomes of the measurement procedure  
[[Borsboom et al., 2004, p. 1061](#)]

As [Borsboom et al. \(2004\)](#) note, an immediate implication of this view of validity (*viz.* that validity requires a real attribute causally influencing assessment procedures) is that the central issue under investigation here is primarily *ontological*—i.e., concerning what the attribute actually is—rather than *epistemological*—i.e., how we come to know the attribute—since coming to know about any attribute requires that the attribute exist in the first place. Additionally, for variations in test performance to reflect changes in the attribute, one further requires an account of process (a theory of response behaviour) to address how the attribute comes to be reflected in test performance. Any theory of response process itself is also necessarily embedded within a broader philosophical position that entails some commitment to an ontological stance (such as determinism) and the logic of explanation (e.g., causes and effects need to be logically distinct—see [Hibberd, 2014](#)). Consequently, since creating valid assessment procedures requires both theory about the nature of the attribute and about how the attribute comes to influence test scores, test validity cannot be outsourced to statistical analyses because such analyses assume valid assessment in the first place: “the problem of validity cannot be solved by psychometric techniques or models alone. On the contrary, it must be addressed by substantive theory. Validity is the one problem in testing that psychology cannot contract out to methodology” ([Borsboom et al., 2004, p. 1062](#)).

Before continuing it should be said that while [Borsboom et al. \(2004\)](#) provide a welcome approach to test validity by addressing the ontological status of the attribute in question (prior to developing assessment procedures), there are further questions concerning the precise nature of quantitative attributes and whether personality attributes are necessarily measurable (see Section 4 of this paper; see also [Michell, 1999, 2009b, 2013](#)). However, for present purposes we turn to a critical assessment of the dispositional trait concept to demonstrate that as the concept stands, there are doubts that any trait-assessment procedures are valid for the simple reason that dispositional traits appear to be fictitious entities.

## 3. Personality traits and test validity

The most commonly assessed personality attribute is that of the personality trait and while there are a variety of trait approaches, traits are generally seen as intra-individual ‘temperament-like variables’ ([Matthews, Deary, & Whiteman, 2003](#); [McCrae & Costa, 1995](#); [McCrae et al., 2000](#)). The identification of traits has been guided by the lexical approach to personality (traceable to Galton, Thurstone, and Cattell, amongst others—[Digman, 1990](#); [Goldberg, 1990](#); [Matthews et al., 2003](#)), utilising trait-term adjectives in language, as well as factor-analytic studies, to develop hierarchical models that identify higher level broad factors from clusters of lower level traits and specific acts ([Eysenck, 1991, 1997](#); [Goldberg, 1990](#); [Matthews et al., 2003](#); [McCrae & Costa, 1997](#)). Traits are commonly considered latent variables that underlie differences in observable test performance—“a trait is not an observable attribute of an individual. Only its behaviour manifestations can be observed” ([Carr & Kingsbury, 1938, p. 509](#); cf. [McCrae & Costa, 1995, 2008](#))—and roughly between three and 16 traits have been identified (see [Cattell, 1957](#); [Eysenck, 1991, 1997](#)).

The Five Factor Theory (FFT) is generally considered to be the dominant trait approach and proposes that five traits (openness to experience, conscientiousness, agreeableness, extraversion, neuroticism) are universal, biologically-based endogenous tendencies or dispositions ([Costa & McCrae, 1997, 1999, 2000, 2008, 2009](#)). As the hypothesised “underlying causes of behaviour” ([McCrae &](#)

Costa, 1995, p. 236; cf. Costa & McCrae, 2008, 2009) these traits, in principle, fulfil the necessary conditions of validity insofar as the attribute is hypothesised both to exist and be causally related to (test) performance. As McCrae and Costa (1995) write, “the causal argument is in principle clear: traits as underlying tendencies cause and thus explain (in general and in part) the consistent patterns of thoughts, feelings, and actions that one sees” (p. 236).

### 3.1. Traits, dispositions and the problem of reification

On the account of test validity discussed here, if the attribute does not exist, then “the test is not valid for measuring that attribute, no matter how useful the test may be for prediction or selection or how well it may fulfil other functions” (Borsboom et al., 2004, p. 1065). Since traits are hypothesised causes of behaviour (including test performance) a first step then in assessing the validity of dispositional trait tests is to examine the coherency of the trait hypothesis using conceptual analysis. Conceptual analysis involves assessing the logical coherency of concepts, theories, and hypotheses, and further contributes to the interpretation of empirical findings, as well as shaping meaningful and valid avenues for future research directions (Bell, Staines, & Michell, 2000; Boag, 2011a, 2011b; Machado & Silva, 2007; Petocz & Newbery, 2010; Wakefield, 2007). Of particular relevance here, conceptual and theoretical work can be used to extricate the logically sound features of theories and reject the incoherent aspects (before attempting to execute empirical tests). As Machado and Silva (2007) write, “[s]cience has always included the screening of concepts and arguments for clarity and coherence” (p. 680), and if a theory fails the logical test, then it can be rejected *a priori*, since an incoherent theory cannot give rise to coherent hypotheses (Petocz & Newbery, 2010). Furthermore, while not always appreciated, any researcher more or less necessarily conducts conceptual research, implicitly or explicitly, since (for example) formulating scales and tests must be guided by some conceptualisation of what one is looking for.

Conceptual analysis is also essential for addressing the first criterion of test validity in a realist account, *viz.* addressing the ontological status of the attribute in question. While there are many possible approaches to conceptual research (see, for instance, Moors & De Houwer, 2006; Wakefield, 2007), a realist approach provides a cogent position for examining the coherency of theories and concepts, in part due to it clarifying the important, but often poorly appreciated, distinction between *qualities* (or properties) and *relations*. As Mackie (1962) writes, “a quality is an intrinsic feature of a thing, it belongs to the thing itself, whereas a relation holds between two or more things” (p. 266), and more specifically, relations involve at least two or more distinct *terms* that must have their own intrinsic properties (to constitute what stands in the relation):

Anything that can stand . . . in any relation at all, must have at least some intrinsic properties. If that were not the case . . . then we could not understand what it was that was said to have those relationships. A relation can only hold between two or more terms, and a part of what is involved in seeing those terms as related is being able to see them as distinct, that is, as each having its own intrinsic properties, so that we can say what the terms *are* that are related. This means that each term of the relation must be able in principle to be described without the need to include any reference to its relation to the other

[Maze, 1983, p. 24, his italics; cf. Maze, 1954, p. 231; Michell, 1988, p. 234]

The first question then is to ask precisely what is meant by a ‘dispositional trait’ because if traits are to stand as an intra-individual attributes causing (test) behaviour then we need to know what

they actually are and how they come to cause variations in test performance (Boag, 2011a, 2011b). The first problem, however, is that concepts such as dispositions and tendencies do not appear to actually refer to specific attributes (i.e., qualities/properties of persons) but instead summarise what a person is likely to do (e.g., persons *S* is disposed to do *y* or *S* has a tendency to do *z*). To say, for instance, that an object has a ‘fragile’ disposition is to say that under certain circumstances (i.e., in relation to other things) the object will *behave* in a particular manner. So if we ask specifically what a disposition or tendency actually is, the answer tends to be in terms of a relation (*S* tends to do *x*) rather than saying anything about the intrinsic features of persons (see Boag, 2011a, 2011b). In terms of the ontological requirements for test validity, the dispositional account thus offers only an implicit reference to what it is that brings such effects about: “to say that glass is fragile is to say that there is something about it which would help causally to bring about its breaking, but not to say what that something is” (Mackie, 1977, p. 364).

Similarly so for personality attributes: for instance, defining the trait of ‘extraversion’ in terms of a ‘tendency’ does not provide any real idea of what it is that is relevant to explaining extraverted behaviour. The tendency, if anything, is a relation between a person *S* and some situation *p* (*SRp*). Since relations cannot be reduced to either term of the relation (either *S* or *p*; see Maze, 1983), traits, so defined, appear precluded from existing as intrinsic properties of persons that could causally influence (test) behaviour. While no one is disputing that there will be relevant properties of a person *S* contributing at least in some way to the response between *S* and situation *p* (for instance, in terms of nervous system properties—Eysenck, 1967, 1997), by defining traits as dispositions (or tendencies), the trait theorist is necessarily committed to the position that traits are greater than *S*’s properties (i.e., traits cannot be reduced to anything less than the relationship between those properties of the individual and situations—*SRp*). Accordingly, traits as ‘tendencies’ and ‘dispositions’ (e.g., McCrae, 2004; McCrae & Costa, 1995, 1999, 2008) fail the logical test since they appear to confuse person–situation relationships with within-subject attributes.

The specific confusion committed here is the logical error of reification. *Reification* involves mistaking relations with properties (Bell, Staines, & Michell, 2001; Boag, 2011a, 2011b; Passmore, 1935), and pertinent here to the trait-concept is confusing what something *does* with what it *is*: consistent patterns of *behaviour* (what a person does) appear to become reified into properties and then used to explain those same behaviours (Boag, 2011a). There is, of course, nothing problematic with hypothesising that people differ on some within-subject variable *X* and that *X* is causally relevant to explaining individual differences in behaviour *Y*. However, here the logic of explanation needs to be kept in mind. If traits are within-subject attributes that causally explain behaviour then traits must also be logically independent of the behaviours that they are said to cause (i.e., causes and effects cannot be conflated), for to say otherwise is to subscribe to circular explanation. Circular explanation occurs when the term used to explain some effect (the *explanans*) is equivalent to the effect that it is said to explain (the *explanandum*), such that the explanation for some occurrence is the occurrence itself (Bell et al., 2001; Boag, 2011a, 2011b). This easily occurs whereby the description of behaviour becomes mistakenly treated as a property of persons and then used to explain that same behaviour (a problem raised at various times in the history of modern psychology—Bandura, 1999; Boag, 2011a, 2011b; Cervone, 1999; Howe, 1990; Kroger & Wood, 1993; Maze, 1954; Skinner, 1953).

To demonstrate the circularity of trait-explanations, consider the problem of ‘verbal-magic’. ‘Verbal-magic’ involves “giving a name to a certain kind of event and then using the name as if it accounted for the *occurrence* of that kind of event” (Maze, 1954, p. 226; his italics). One problem with verbal magic is that we have

only the illusion of knowing the relevant causes when in fact we have no clear idea of what the actual causes are (cf. Boag, 2011a; Cervone, 1999). To illustrate this, consider Terracciano and McCrae's (2012) recent defence of the causal role of traits by analogy with explaining the flocking behaviour of birds via the concept of a 'flocking instinct':

Presumably some species flock because they have evolved a mechanism, or suite of mechanisms, that allow each bird to respond to its environment in ways that give rise to flocking behaviour. Loosely, we can call this set of mechanisms a *flocking instinct*, present in some species but not others. Surely it is meaningful to say that the innate flocking instinct is a *cause* of flocking behaviour, because when it is absent, flocking does not occur

[Terracciano and McCrae, 2012, p. 449, their italics]

However, here Terracciano and McCrae commit the same spurious reasoning as early 20th century instinct theorists such as McDougall (1923) who believed that instincts could be identified from observing behavioural goals (pp. 118–119). The specific problem here is that attributing an instinct to explain any given behaviour is vacuous since we have then as many instincts as there are recognisable acts of behaviour, as well as no behaviour-independent evidence for the supposed instinct. Holt, commenting on McDougall's instinct theory in the 1930s, recognised this specious explanatory strategy:

... man is impelled to action, it is said, by his instincts. If he goes with his fellows, it is "herd instinct" which activates him; if he walks alone, it is the "anti-social instinct"; if he fights, it is the "pugnacity instinct"; if he defers to another it is the instinct of "self-abasement"; if he twiddles his thumbs, it is the thumb-twiddling instinct; if he does not twiddle his thumbs, it is the thumb-not-twiddling instinct. Thus, everything is explained with the facility of magic—word magic

[Holt in Yankelovich, 1973, p. 413]

Cramer et al., in reply to Terracciano and McCrae, similarly observe that any appeal to a 'flocking instinct' is comparable to saying that since "some women prefer high heels whereas others do not ... it must be so that an underlying *instinct to wear high heels* exists that causes these behavioural differences between women" (Cramer et al., 2012b, p. 454, their italics). In other words, to attempt to explain any given behaviour by proposing an instinct (or faculty, etc.) responsible for it is simply vacuous since the 'evidence' for the instinct is always available:

Drives [instincts] specified by aim can be postulated without check, because the 'evidence' for them is always available: the observed behaviour they were postulated to explain. Any commonly occurring behaviour can be 'explained' by saying there must be an instinct or drive behind it, but it is only a pseudo-explanation.

[Maze, 1993, pp. 462–463; cf. Freud, 1915, pp. 123–124.]

Thus, if traits are identified solely from what people do (or say that they do), and subsequently used to explain the same behaviour that they are inferred from, then traits become the thumb-twiddling instincts of the 21st century. What is instead required is an account of what traits actually are, rather than what they are said to do. If there is no independent evidence for dispositional traits apart from the behaviour that traits are said to explain then we have simply again "an instance of word-magic, an instance of the fallacy of reification, inventing entities in an *ad hoc* fashion to do explanatory jobs. Such entities have no qualitative nature of their own; their sole existence lies in observed relations of certain kinds ..." (McMullen, 1982, p. 224).

Let us then return to the question of the ontology of traits since this is essential for considering test validity. A variety of causal dynamics could potentially rescue the trait concept from the logical problems discussed above (see, for instance, Gerber, 2011). Perhaps the trait concept could be salvaged in terms of biological attributes (along, say, the lines of Eysenck's (1967, 1997) theory). However, McCrae and Costa reject characterising traits as biological attributes since traits are 'psychological structures' (McCrae & Costa, 1995, 1999): "traits are ... real psychological structures" (McCrae, 2004, p. 4) and "ours is a psychological theory of personality. We do not equate basic dispositions with biological constructs, nor have we offered a psychobiological theory of personality ..." (McCrae & Costa, 1995, p. 239). Perhaps, then, traits refer to constellations of desires and beliefs contributing to consistencies in behaviour? McCrae and Costa however explicitly reject this suggestion, instead classifying desires and beliefs as *effects* of traits ('characteristic adaptations') (McCrae & Costa, 1995, p. 230; McCrae, 2004, p. 5). So what, then, are the features of these psychological traits? Well, according to McCrae and Costa we can never know: "traits ... are directly accessible neither to public observation nor to private introspection. Instead, they are deeper psychological entities that can only be *inferred* from behaviour and experience" (McCrae & Costa, 1999, p. 143, their italics). That is, traits are immune to empirical criticism, comparable to other hypothesised 'variables' such as the 'soul' (see Boag, 2011a, pp. 234–235) and so even if traits exist we could never know anything about such attributes to know that we are validly assessing them.

The conclusions above are hardly novel and will not be news to trait-theorists themselves (e.g., McCrae & Costa, 2008) and have been discussed similarly in connection to intelligence as an explanatory variable (e.g., Gould, 1981; Howe, 1990; Passmore, 1935). In many respects a pressing question for psychology is why reification persists (Boag, 2011a). However, there is of course nothing to prevent us from defining traits as person–situation *relations* (cf. Mischel & Shoda's (1995) *if ... then ...* situation–behaviour signatures). Recognising traits as person–situation relationships would mean that we can begin asking appropriate questions concerning the terms of the relation and the actual causal dynamics underlying those relationships. For example, if the processes mediating *if ... then ...* person–situation relationships end up entailing combinations of cognitive–affective units (Mischel, 2004; Mischel & Shoda, 1995), motivational drives and affective systems (Maze, 1983; Panksepp, 2005) and/or various evolutionary–adaptive processes (Marsh & Boag, 2013) (etc.) then we are a step closer to providing a logically coherent explanatory account of trait–relations that could help provide an understanding of the relationship between attributes, response process, and test performance, which is necessary for valid tests of personality.

### 3.2. *Misinterpreting factor analysis as a source of the problem*

The problems of reification and circular explanation appears to arise (at least, in part) through factor analytic research which identifies behaviourally-recurrent trends in populations, and then mistakenly uses these trends to identify intrinsic features of persons causing those same item responses (Marsh & Boag, 2013; see also discussion concerning confusing between-subject variability with within-subject variables—Borsboom, Mellenbergh, & van Heerden, 2003; Cervone, 2005; Lamiell, 2007, 2013; Vautier, 2011; Vautier, Lacot, & Veldhuis, 2014; Vautier, Veldhuis, Lacot, & Matton, 2012). In this respect the limits of factor analysis are often not carefully considered (Guilford, 1975, p. 802; cf. Matthews et al., 2003), although it is also worth noting that many early factor analytic researchers were well aware of the potential of reifying false dispositions, such as inferring that someone has an 'irritating disposition' because s/he is annoying to others (e.g., Carr & Kingsbury, 1938;

Thomson, 1951, p. 59; see also Anastasi, 1961, p. 347). Note, too, that identifying factors easily oversimplifies the situation by suggesting that a *single* attribute (e.g., extraversion) explains behaviour (e.g., sociability) rather than appreciating the complexity of causes, effects and causal field (cf. Cramer et al., 2012a, p. 417; Nilsson, 2014, p. 21).

Furthermore, in terms of test validity, factor analysis gets it the wrong way round: the test scores employed within factor analyses assume that the tests are valid in the first place, but of course, if they are not, then how can validity suddenly emerge from invalidity?

What is constitutive of validity is the existence of an attribute and its causal impact on scores. Therefore, if one does not have an idea of how the attribute variations produce variations in measurement outcomes, one cannot have a clue as to whether the test measures what it should measure. No table of correlations, no matter how big, can be a substitute for knowledge of the processes that lead to item responses. The knowledge of such processes must be given by substantive psychological theory and cannot be based on methodological principles.

[Borsboom et al., 2004, p. 1068]

Consequently, the top-down factor analytic approach is not a valid procedure for understanding either the nature of the attribute in question or the causal mechanism pertinent to understanding precisely how the attribute influences test scores. Instead, rather than jumping to producing assessment procedures to help identify traits (devising procedures, collecting data, and using factor analysis to deduce attributes), one needs to *first* develop a coherent theory of traits in order to develop valid assessment procedures, a point observed by Boring as early as the 1920s:

The more we know of the intimate nature of the entity with which we are dealing the more accurate and complete can our descriptions become. But, if in psychology we must deal—and it seems we must—with abilities, capacities, dispositions and tendencies, the nature of which we cannot accurately define, then it is senseless to seek in the logical process of mathematical elaboration a psychologically significant precision that was not present in the psychological setting of the problem. Just as ignorance will not breed knowledge, so inaccuracy of definition will never yield precision of result.

[Boring, 1920, p. 33; cf. Cattell, 1943; Michell, 2009a]

The basic upshot from this discussion—and it really is a fairly obvious point—is that developing valid assessment of any attribute requires a coherent theory of the attribute prior to developing assessment procedures. Conversely, if a researcher is confused about an attribute (and worse, if s/he does not know that s/he is confused) then it is difficult to see how appropriate tests of the attribute could ever be developed. Simply put, how can anyone devise a test of *X* without clearly knowing what *X* is? Accordingly there is an essential role of theory to guide personality assessment by providing a more precise understanding of what is actually being talked about prior to developing tests (Boag, 2011a, 2011b; Borsboom et al., 2004, 2009; Petocz & Newbery, 2010).

#### 4. Valid tests and the theory of measurement

The preceding discussion indicates that we are simply not in a position to assess the test validity of dispositional trait assessment procedures simply because, at best, the trait concept is too vague, and at worst, traits are fictitious attributes derived from reifying behaviour. However, there is a further problematic assumption in much personality assessment concerning whether we are entitled

to claim to be *measuring* personality attributes in any serious meaning of the word. Consider the way that personality tests are commonly constructed: a number of items are developed and then individuals either respond to behavioural descriptions (e.g., *I tend to avoid parties*) or judge others (e.g., *to what degree is John fearful?*) (Digman, 1990) and the collected data are then reduced via factor analysis to a smaller set of factors in order to understand the trait attribute (see Goldberg, 1990). Within this process, numerical values are assigned to the trait (the 'theoretical' variable) and thus a *representational* theory of measurement is typically employed. The representational theory of measurement proposes that measurement "is the assignment of numerals to objects or events according to rule" (Stevens in Mari, 2005, p. 263; cf. Stevens, 1946) and is the common approach to measurement in personality research (e.g., Haslam, 2007). Numerical values are commonly assigned to hypothesised personality attributes via Likert-style response formats (i.e., numerical rating responses on a scale of strongly disagree = 1 to strongly agree = 5). This assignment of numerals is then typically 'treated' as interval scale of measurement (where the intervals are believed to be equidistant), even if properly considered nominal or ordinal data. While seemingly irrational, this approach nevertheless was sanctioned by Lord et al. (1968) who explicitly "treat . . . measurement as having an interval scale properties, although it is clear that the measurement procedure and the theory underlying it yield only a nominal or, at best, an ordinal Scale" (p. 22), a view that has become subsequent orthodoxy: "the primary advantage of rating-scale questions is that they produce numerical values that *can be treated as measurements* from an interval scale. (Recall . . . that an interval scale consists of a series of equal-sized categories, which make it possible to measure distances on the scale.)" (Gravetter & Forzano, 2012, p. 378, italics added; see also Haslam, 2007, p. 183).

On the realist account of test validity this is obviously problematic because there is no consideration of the actual structure of the attribute in question—it is simply treated as a quantitative attribute—and the relation between the attribute (assuming that the attribute even exists) and test scores is unknown. On the representational view of measurement, one could, presumably, develop a scale 'measuring' the thumb-twiddling trait using the above approach (via ratings on items such *I tend to twiddle my thumbs*). It is also clear that we do not treat all and any assignment of numerals as measurement. For instance, we generally do not consider the numerical values assigned to undergraduate essays and reports to constitute 'measurement' in any substantive sense of the word. And yet, the process is identical to what is occurring in personality assessment (the assignment of numerals according to rule). A question then is whether the assignment of numerals, as typically found in personality assessment, is really measurement, especially given the earlier analysis whereby we appear to be 'measuring' fictitious entities.

To address this, consider the definition of measurement as found in the so-called hard sciences whereby measurement is "*the estimation or discovery of the ratio of some magnitude of a quantitative attribute to a unit of the same attribute*" (Michell, 1997, p. 358, his italics). On this view of measurement:

. . . it is theorized that an attribute, such as length, has a distinctive kind of internal structure, viz., *quantitative structure*. Attributes having this kind of structure are called *quantities*. Following a well-established usage, specific instances of a quantity are called *magnitudes* of that quantity (e.g., the length of this page is a magnitude of the quantity, length). Magnitudes of a quantity are measurable because, in virtue of quantitative structure, they stand in relations (*ratios*) to one another that can be expressed as real numbers

[Michell, 1997, p. 356, his italics]

Here it is the nature of the attribute that allows measurement rather than simply treating the assignment of numerals as sufficient. Furthermore, on the above view, one can sensibly speak of *units* of measurement (as found in true interval and ratio scales), since each unit is equidistant and these units can be sensibly added together to measure the actual quantity of the attribute. For instance, with length, the interval between 1 and 2 cm is equal to the interval between 99 and 100 cm and these can be summed to know actual amounts (unlike rank-orders where the distance between rankings is unknown, and so it makes no sense to add the rankings together). What allows such true units of measurement (and thus for the attribute to be quantitative and hence possibly measurable, strictly speaking) is that the attribute possesses both *ordinal* and *additive* structure (see [Michell, 1999, Chap. 3, 2009a](#)). Whether any attribute possesses such structure requires investigation and while an attribute might display order this alone is not sufficient for assuming quantity since additivity must also be satisfied (see [Michell \(1999, 2005, 2008, 2009a, 2009b\)](#) for further discussion of quantitative structures). For instance, Michell provides the example of ‘functional ability’ which appears to be an ordinal, non-quantitative attribute: elderly people generally lose various abilities in a particular order (the ability to independently climb stairs is lost before getting into a bathtub unaided, for instance). Nevertheless, while order is present, and we can speak of various *levels* of independence, the differences between abilities (say, climbing stairs and getting into the bath) appear to be qualitative rather than quantitative ones and it would not make sense to add these differences together (see [Michell, 2009b](#) for further discussion).

In personality research we typically only observe that responses are ordered. We might observe that people are more or less extraverted from one another, and from the respondent’s point of view, his or her responses (on, say, a Likert-scale) reflect either more or less modal agreement with the item. Consequently, we are at best dealing with ordinal rankings ([Michell, 2003, 2005; Stevens, 1946](#)) and there is no reason to assume that any personality attribute (such as ‘extraversion’, presuming that such an attribute even exists) also sustains additive structure to allow true measurement. Of course, the personality tests themselves easily allow the mistaken belief that the scale is assessing true ‘units’ of personality measurement. Visually, for instance, a Likert response format appears like an interval scale and Likert himself assumed that his data reflect scores where “the units of which are equal throughout the entire range” ([Likert, 1932, p. 42](#)). However, since there is no clear connection between the numerical values and how they map on to the attribute in question, and respondents simply rate ‘more’ or ‘less’ in relation to items, what we have at best is ordinal rankings (rank-order, where relative differences amongst values are not known to be equal) rather than interval or ratio scales required for measurement *per se* (again, not a novel claim; see [Göb, McCollin, & Ramalhoto, 2007; Jamieson, 2004; Kuzon, Urbanchek, & McCabe, 1996](#)). The use of numerals in personality research is thus highly suspect. As [Jamieson \(2004\)](#), paraphrasing [Kuzon et al. \(1996\)](#) writes, “the average of ‘fair’ and ‘good’ is not ‘fair-and-a-half’; this is true even when one assigns integers to represent ‘fair’ and ‘good’!” (p. 1218).

Part of the problem here is that researchers generally do not have an adequate theory of measurement and operate via the mistaken belief that we “can make something quantitative simply by throwing numerals at it” ([Petocz & Newbery, 2010, p. 133](#)). Believing that assigning numerals is sufficient for constituting measurement is a myth based on convenience or desire rather than good science ([Michell, 2000, 2008, 2009a, 2009b](#)). As [Michell \(2009a\)](#) writes, “[u]nits presume quantity and in science the hypothesis that attributes are quantitative, like any empirical hypothesis, is not made true by wishing” (p. 114). In some respects, the representational theory of measurement represents Freudian

wish-fulfilment (desiring that *X* be the case, believing *X*, while ignoring the real situation) ([Boag, 2015](#)).

## 5. The way forward: the significance of theory

Presently there are indications that personality trait research is not only chasing fictitious reified entities (traits) but that such mistaken entities are then assumed to be quantitative attributes when at best test scores provide ordinal rankings of modal responses. So how do we go forward? If we hypothesise that a personality attribute, *X*, exists and further wish to then develop valid procedures for assessing (and potentially measuring) *X*, then the obvious starting point is to *first* develop an explicit theory of the attribute (*X*) (see also [Sijtsma, 2012](#)). Developing explicit theory here involves conceptual analysis of the theory of the attribute and how the attribute might causally relate to any proposed assessment procedures. This initially involves assessing the logical coherency of any theory using conceptual analysis, prior to submitting the theory to observational tests ([Boag, 2011b; Michell, 2000; Petocz & Newbery, 2010](#)). If the theory fails the logical test, then it can be rejected *a priori* (or at least modified to address shortcomings), since an incoherent theory cannot give rise to coherent hypotheses. Any theory of an attribute is also necessarily embedded within a greater theoretical system (including commitments to causality and understanding the logic of explanation—[Hibberd, 2014](#)) and so psychological research necessarily has philosophical underpinnings to consider. Further spelling out how the particular attribute involved stands *vis-à-vis* other attributes is also necessary for addressing the process of how any attribute will be reflected in test scores.

Another way of considering this is to take into account Michell’s distinction between the *scientific* and *instrumental* tasks in science ([Michell, 1997, 2000, 2009b](#)). If personality researchers are interested in scientific assessment and (potentially) measuring personality attributes such as traits then the actual structure of the attribute needs to be examined before assuming that the attribute is quantitative (the scientific task). Addressing whether any attribute possesses both ordinal and additive structure requires both a theory of the attribute and conducting (direct or indirect) empirical tests before asserting that such attributes are quantitative ([Michell, 1997, 1999, 2009a](#)). This approach requires theoretical and conceptual analysis (throughout the entire process—[Petocz & Newbery, 2010](#)) as well as empirical investigation of the structures in question (tests of whether the attribute is in fact quantitative). Conceptual and empirical investigation can possibly reveal a number or known attribute structures including classificatory structures (e.g., nationality), a range of various ordinal structures (partial, weak, or simple orders), as well as quantitative structure of continuous attributes ([Michell, 1997, 1999, 2009a](#)). In the latter case:

If you are going to seriously test the hypothesis that some latent trait, *X*, is quantitative, then *X* must be specified in sufficient detail for its hypothesized quantitative structure to have a theoretical interpretation in terms of item structures and the psychological processes

[[Michell, 2008, p. 15](#); cf. [Boring, 1920](#)]

Once the attribute’s *actual* structure is known then the researcher is in the position of knowing which procedures will allow validly assessing this attribute ([Petocz & Newbery, 2010](#)). The instrumental task here involves developing test items that are sensitive to the properties of the attribute under investigation. Without adhering here to the logic of scientific investigation we run the risk of self-deceiving ourselves about the scientific merit of our investigation ([Michell, 1997, p. 359](#)).

Keeping in mind that as scientists we are trying to understand the way the world is, rather than as we wish it to be, there is no shame should an attribute be found to be non-quantitative since this is the true state of affairs to be discovered. The view that measurement is a *sine qua non* of science (the quantitative imperative—[Michell, 2003](#)) is in fact dogmatic and anti-scientific (see also [Petocz & Newbery, 2010](#)). Should any personality attribute be found to be, say, simply ordinal rather than quantitative then—as the true state of affairs—this is open to scientific enquiry like anything else that actually exists. As Michell notes “our primary goal in science is not to *presume* answers to questions, but to *discover* the real structure of attributes” ([Michell, 2009a, p. 118](#), his italics). Obviously, then, personality researchers can also embrace qualitative methods *where appropriate* ([Michell, 2003, 2004](#); [Petocz & Newbery, 2010](#)): “A scientifically healthy psychology will embrace qualitative methods as enthusiastically as quantitative ones and devote substantial energies to investigating the quantity–quality distinction as an empirical issue” ([Michell, 2003, p. 25](#)). On the other hand, if warranted one can then hypothesise that a given attribute is quantitative and go on to test whether this is in fact the case. Conjoint measurement (involving a hierarchy of cancellation conditions) is one possible means available for assessing whether a variable is merely ordinal or quantitative ([Michell, 2005](#)). However, Item Response Theory (IRT), which “assumes item scores to be a function of an underlying latent variable” ([Borsboom & Mellenbergh, 2004, p. 108](#)), appears problematic for assessing quantitative structure given that the nature of the attribute is not sufficiently addressed *prior* to data collection (see also [Michell, 2004](#); [Petocz & Newbery, 2010](#); [Sijtsma, 2012](#); [Vautier et al., 2012](#) for further problems associated with Item Response Theory).

Unfortunately, it is doubtful whether greater attention to theory and the logic of measurement is likely to appeal to the mainstream personality researcher, simply due to the need to rapidly accumulate data and disseminate findings (i.e., meet the publish or perish demands of academia). Perhaps conceding this, [Barrett \(2003\)](#) proposes a pragmatic direction that he terms ‘applied numerics’ whereby researchers continue with their current practices “for the purpose of approximating loose theoretical or pragmatic hypotheses” (p. 433), while acknowledging that validity is nevertheless compromised.

... if the process of mapping numbers onto psychological attributes is recognised from the outset as an approximation *with no great regard paid to the scientific value* of such an enterprise, then this constitutes an honest approach that has indeed paid many pragmatic dividends. As the history of applied psychometrics has demonstrated, many variables have been constructed and utilised as predictive indicators of practically relevant phenomena (such as job satisfaction, employee wellbeing, personality, IQ), without any explicit theory of the meaning of the variables other than a “common-sense” meaning that is generally applied to assist in their interpretation... This is not a “scientific” approach, but rather, a pragmatic approach

[[Barrett, 2003, p. 433](#), italics added]

In many respects, this would mean business as usual with the addition of recognising that the validity of one’s testing is compromised. The problem with this, however, is that no one applying for research funding (for instance) is likely to ever admit to conducting research with “no great regard paid to the scientific value” of the enterprise. Furthermore, ‘applied numerics’ will simply perpetuate current areas of confusion in psychology research. For example, given the supposed replicability crisis in psychology (see [Pashler & Wagenmakers, 2012](#), and related articles), finding consistent replicable results is not going to be helped by testing “loose

theoretical or pragmatic hypotheses” via inappropriate statistical analyses and inappropriately considered data. Instead, a much greater consideration of theory is needed for a true science of personality: both of personality generally and assessment/measurement more specifically. As mentioned above, the only logical place to begin would be to first clarify what we are talking about prior to developing assessment procedures. If a personality researcher wishes to develop an assessment procedure for some personality attribute *X* then s/he needs first to develop a logically coherent theory of both *X* is and how *X* might be causally related to scores on the proposed assessment procedure.

[Eysenck \(1997\)](#) once noted that we “deal with persons, not atoms” (p. 1234) and it should be kept in mind that personality theory itself provides the greatest guide for developing valid personality assessment procedures when we consider the ‘whole person’. The reason for this is simply that understanding the various components of personality and their interrelationships is more likely to provide a more comprehensive understanding of the causal processes that lead to test performance. For instance, motivational and affective factors will be involved in any test performance and thus personality assessment should consider such processes in any theory of response behaviour. Accordingly, rather than attempting then to understand isolated fragments, personality assessment will benefit from greater consideration of a more integrated view of persons (an issue raised in different ways by various authors—[Cervone, 2005](#); [Mayer, 2005](#); [McAdams & Olson, 2010](#); [Mischel, 2004](#); [Nilsson, 2014](#)). By addressing a general theory of persons—a *metapsychology*—we might then begin to appreciate the necessary and sufficient conditions for explaining a person’s behaviour (necessary for a theory of response behaviour) and avoid overly-simplistic thumb-twiddling (trait/instinct/faculty) explanatory strategies as is currently observed in trait approaches. Such an integrative approach will entail not only considering quality and quantity, brain and mind, and the nature of relations including causality, but also an understanding of how specific attributes relate to other facets of personality such as motivation, affects, and beliefs. Understanding these attributes requires preliminary definitional work and thus ‘meaning’ is a core concern for empirical research (cf. [Maraun & Gabriel, 2013](#); [Petocz & Newbery, 2010](#)).

A final caveat should be to note that the discussion here is chiefly concerned with dispositional traits such as found in the FFT/FFM. There are many approaches to personality that are explicitly committed to theory, employing a more or less theory-driven, bottom-up approach to personality assessment. For example, both Reinforcement Sensitivity Theory (and its revisions—e.g., [Jackson, Levine, & Furnham, 2003](#); [Smillie, Pickering, & Jackson, 2006](#)) and affective neuroscience approaches ([Davis, Panksepp, & Normansell, 2003](#)) provide theoretically driven approaches to scale content development. However, while attention to personality theory is naturally welcome, it is only a first step towards valid testing since not only is a theory of the attribute(s) in question required, but also a theory of measurement, should the researcher wish to assess quantities. Nevertheless, such approaches have much greater potential for providing the requisite basis needed for a scientifically healthy approach to personality assessment.

## 6. Conclusion

Developing valid assessment of any attribute requires a coherent theory of the attribute prior to developing assessment procedures. We are presently not in a position to assess the test validity of dispositional trait assessment procedures simply because, at best, the trait concept is too vague, and at worst, traits are fictitious attributes derived from reifying behaviour. In personality research it would be

fair to say that too little attention has been paid to the entire scientific enterprise with respect to the scientific examination of personality attributes such as traits, either theoretically (whether traits are conceptually coherent) or empirically (whether traits, should they survive the logical test, have a quantitative structure). The assumption that personality attributes are quantitative and that a 'scale of measurement' can be imposed upon them, as in the case with Likert-type response formats, undermines any claim to validity since the actual structure of the attribute is ignored. If we are to take the science of personality seriously then such issues need to be addressed. Recognising the possibility of error and paying special attention to the means for both recognising and preventing error is what makes science superior as a method of enquiry (Michell, 2000, 2008; Petocz and Newbery, 2010), and given that it is clearly possible to *misconceptualise* a person's characteristics based on observable behaviours (e.g., classifying individuals as 'witches'), there is an obvious need to use whatever means available to us as scientists to avoid making mistakes. Empirical research is one tool available but there are also particular theoretical issues which hinder personality assessment and which are not simply absolved through further empirical research alone (Boag, 2011b; Petocz and Newbery, 2010). While conceptual and theoretical precision necessitates an explicit commitment to considering 'philosophical' issues, such an undertaking should not be beyond the capabilities of any serious scientific researcher.

## References

- Anastasi, A. (1961). *Psychological testing*. New York: Macmillan.
- Bandura, A. (1999). Social cognitive theory of personality. In D. Cervone & Y. Shoda (Eds.), *The coherence of personality: Social-cognitive bases of consistency, variability, and organization* (pp. 185–241). New York: Guilford Press.
- Barrett, P. (2003). Beyond psychometrics: Measurement, non-quantitative structure, and applied numerics. *Journal of Managerial Psychology*, 18, 421–439.
- Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, 9, 9–30.
- Bell, P., Staines, P., & Michell, J. (2000). *Logical psych*. Sydney: UNSW Press.
- Berry, C. M., Ones, D. S., & Sackett, P. R. (2007). Interpersonal deviance, organizational deviance, and their common correlates: A review and meta-analysis. *Journal of Applied Psychology*, 92, 410–424.
- Boag, S. (2011a). Explanation in personality research: 'Verbal magic' and the Five-Factor Model. *Philosophical Psychology*, 24, 223–243.
- Boag, S. (2011b). The role of conceptual analysis in personality research. In S. Boag & N. Tiliopoulos (Eds.), *Personality and Individual Differences: Theory, Assessment and Application* (pp. 321–330). New York: Nova.
- Boag, S. (2015). Repression, defence, and the psychology of science. In S. Boag, L. A. W. Brakel, & V. Talvitie (Eds.), *Philosophy, science and psychoanalysis*. London: Karnac.
- Boring, E. G. (1920). The logic of the normal law of error in mental measurement. *The American Journal of Psychology*, 31, 1–33.
- Borsboom, D., Cramer, A. O., Kievit, R. A., Zand Scholten, A., & Franic, S. (2009). The end of construct validity. In R. W. Lissitz (Ed.), *The concept of validity: Revisions, new directions, and applications* (pp. 135–170). Charlotte, NC: Information Age Publishing.
- Borsboom, D., & Mellenbergh, G. J. (2004). Why psychometrics is not pathological: A comment on Michell. *Theory & Psychology*, 14(1), 105–120.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, 110, 203–219.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, 111, 1061–1071.
- Carr, H. A., & Kingsbury, F. A. (1938). The concept of traits. *Psychological Review*, 45, 497–525.
- Cattell, R. B. (1943). The description of personality. I. Foundations of trait measurement. *Psychological Review*, 50, 559–594.
- Cattell, R. B. (1957). *Personality and motivation structure and measurement*. Oxford, UK: World Book Co.
- Cervone, D. (2004). The architecture of personality. *Psychological Review*, 111, 183–204.
- Cervone, D. (2005). Personality architecture: Within-person structures and processes. *Annual Review of Psychology*, 56, 423–452.
- Cervone, D. (1999). Bottom-up explanation in personality psychology: The case of cross-situational coherence. In D. Cervone & Yuichi Shoda (Eds.), *The coherence of personality: Social-cognitive bases of consistency, variability, and organization* (pp. 303–341). New York: Guilford Press.
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7, 309–319.
- Costa, P. T., & McCrae, R. R. (2009). The five-factor model and the NEO inventories. In J. N. Butcher (Ed.), *Oxford handbook of personality assessment* (pp. 299–322). Oxford: Oxford University Press.
- Cramer, A. O., Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S. H., et al. (2012a). Dimensions of normal personality as networks in search of equilibrium: You can't like parties if you don't like people. *European Journal of Personality*, 26, 414–431.
- Cramer, A. O., Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S. H., et al. (2012b). Measurable like temperature or mereological like flocking? On the nature of personality traits. *European Journal of Personality*, 26, 451–459.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281–302.
- Davis, K. L., Panksepp, J., & Normansell, L. (2003). The affective neuroscience personality scales: Normative data and implications. *Neuropsychanalysis*, 5, 57–69.
- Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41, 417–440.
- Eysenck, H. J. (1967). *The biological basis of personality*. Springfield: Charles C. Thomas Publisher.
- Eysenck, H. J. (1991). Dimensions of personality: 16, 5 or 3?—Criteria for a taxonomic paradigm. *Personality and Individual Differences*, 12, 773–790.
- Eysenck, H. J. (1997). Personality and experimental psychology: The unification of psychology and the possibility of a paradigm. *Journal of Personality and Social Psychology*, 73, 1224–1237.
- Freud, S. (1915). *Instincts and their vicissitudes*. Standard edition (Vol. XIV). London: Hogarth.
- Furnham, A., & Fudge, C. (2008). The Five Factor model of personality and sales performance. *Journal of Individual Differences*, 29, 11–16.
- Gerber, J. (2011). Six solutions to the circular nature of trait psychology. In S. Boag & N. Tiliopoulos (Eds.), *Personality and individual differences: Theory, assessment and application* (pp. 297–306). New York: Nova.
- Göb, R., McCollin, C., & Ramalhoto, M. F. (2007). Ordinal methodology in the analysis of Likert scales. *Quality & Quantity*, 41, 601–626.
- Goldberg, L. R. (1990). An alternative "description of personality": The big-five factor structure. *Journal of Personality and Social Psychology*, 59, 1216–1229.
- Gould, S. (1981). *The mismeasure of man*. New York: W. W. Norton & Co.
- Gravetter, F. J., & Forzano, L.-A. B. (2012). *Research methods for the behavioral sciences* (4th ed.). Belmont, CA: Wadsworth.
- Guilford, J. P. (1975). Factors and factors of personality. *Psychological Bulletin*, 82, 802–814.
- Haig, B. D. (2014). *Investigating the psychological world: Scientific method in the behavioral sciences*. Cambridge, MA: MIT Press.
- Haslam, N. (2007). *Introduction to personality and intelligence*. Los Angeles: Sage.
- Hibberd, F. J. (2014). The metaphysical basis of a process psychology. *Journal of Theoretical and Philosophical Psychology*, 34, 161–186.
- Hood, S. B. (2009). Validity in psychological testing and scientific realism. *Theory & Psychology*, 19, 451–473.
- Hood, S. B. (2013). Psychological measurement and methodological realism. *Erkenntnis*, 78, 739–761.
- Howe, M. J. A. (1990). Does intelligence exist? *The Psychologist*, 3, 490–493.
- Jackson, C. J., Levine, S. Z., & Furnham, A. (2003). Gray's model of personality and aggregate level factor analysis. *European Journal of Personality*, 17, 397–411.
- Jamieson, S. (2004). Likert scales: How to (ab) use them. *Medical Education*, 38, 1217–1218.
- Kroger, R. O., & Wood, L. A. (1993). Reification, "faking", and the Big Five. *American Psychologist*, 48, 1297–1298.
- Kuzon, W. M., Jr, Urbanchek, M. G., & McCabe, S. (1996). The seven deadly sins of statistical analysis. *Annals of Plastic Surgery*, 37, 265–272.
- Lamiell, J. T. (2007). On sustaining critical discourse with mainstream personality investigators problems and prospects. *Theory & Psychology*, 17, 169–185.
- Lamiell, J. T. (2013). Statisticism in personality psychologists' use of trait constructs: What is it? How was it contracted? Is there a cure? *New Ideas in Psychology*, 31, 65–71.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22.
- Lord, F. M., Novick, M. R., & Birnbaum, A. (1968). *Statistical theories of mental test scores*. Reading, Massachusetts: Addison-Wesley Publ.
- Machado, A., & Silva, F. J. (2007). Toward a richer view of the scientific method: The role of conceptual analysis. *American Psychologist*, 62, 671–681.
- Mackie, J. L. (1962). The philosophy of John Anderson. *Australasian Journal of Philosophy*, 40, 265–282.
- Mackie, J. L. (1977). Dispositions, grounds, and causes. *Synthese*, 34, 361–370.
- Maraun, M. D., & Gabriel, S. M. (2013). Illegitimate concept equating in the partial fusion of construct validation theory and latent variable modeling. *New Ideas in Psychology*, 31, 32–42.
- Mari, L. (2005). The problem of foundations of measurement. *Measurement*, 38, 259–266.
- Marsh, T., & Boag, S. (2013). Evolutionary and differential psychology: Conceptual conflicts and the path to integration. *Frontiers in Psychology*, 4, 1–15.
- Matthews, G., Deary, I. J., & Whiteman, M. C. (2003). *Personality traits*. Cambridge: Cambridge University Press (initially published in 1998).
- Mayer, J. D. (2005). A tale of two visions: Can a new view of personality help integrate psychology? *American Psychologist*, 60, 294–307.
- Maze, J. R. (1954). Do intervening variables intervene? *Psychological Review*, 61, 226–234.
- Maze, J. R. (1983). *The meaning of behaviour*. London: Allen & Unwin.

- Maze, J. R. (1993). The complementarity of object-relations and instinct theory. *International Journal of Psycho-analysis*, 74, 459–470.
- McAdams, D. P., & Olson, B. D. (2010). Personality development: Continuity and change over the life course. *Annual Review of Psychology*, 61, 517–542.
- McCrae, R. R. (2004). Human nature and culture: A trait perspective. *Journal of Research in Personality*, 38, 3–14.
- McCrae, R. R., & Costa, P. T. Jr., (1995). Trait explanations in personality psychology. *European Journal of Personality*, 9, 231–252.
- McCrae, R. R., & Costa, P. T. Jr., (1997). Personality trait structure as a human universal. *American Psychologist*, 52, 509–516.
- McCrae, R. R., & Costa, P. T. Jr. (2008). Empirical and theoretical status of the five-factor model of personality traits. In G. Boyle, G. Matthews, & D. Dacklofske (Eds.), *Sage handbook of personality theory and assessment* (pp. 273–294). Los Angeles: Sage.
- McCrae, R. R., Costa, P. T., Jr, Ostendorf, F., Angleitner, A., Hřebíčková, M., Avia, M. D., et al. (2000). Nature over nurture: Temperament, personality, and life span development. *Journal of Personality and Social Psychology*, 78, 173–186.
- McCrae, R. R., & Costa, P. T. Jr., (1999). A five-factor theory of personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 139–153). New York: Guilford Press.
- McDougall, W. (1923). *An outline of psychology*. London: Methuen & Co.
- McMullen, T. (1982). A critique of humanistic psychology. *Australian Journal of Psychology*, 34, 221–229.
- Michell, J. (1988). Maze's direct realism and the character of cognition. *Australian Journal of Psychology*, 40, 227–249.
- Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology*, 88, 355–383.
- Michell, J. (1999). *Measurement in psychology*. Cambridge University Press.
- Michell, J. (2000). Normal science, pathological science and psychometrics. *Theory and Psychology*, 10, 639–667.
- Michell, J. (2003). The quantitative imperative positivism, naive realism and the place of qualitative methods in psychology. *Theory & Psychology*, 13, 5–31.
- Michell, J. (2004). The place of qualitative research in psychology. *Qualitative Research in Psychology*, 1, 307–319.
- Michell, J. (2005). The logic of measurement: A realist overview. *Measurement*, 38, 285–294.
- Michell, J. (2008). Is psychometrics pathological science? *Measurement*, 6, 7–24.
- Michell, J. (2009a). The psychometricians' fallacy: Too clever by half? *British Journal of Mathematical & Statistical Psychology*, 62, 41–55.
- Michell, J. (2013). Constructs, inferences, and mental measurement. *New Ideas in Psychology*, 31, 13–21.
- Michell, J. (2009b). Invalidity in validity. In R. W. Lissitz (Ed.), *The concept of validity: Revisions, new directions and applications* (pp. 111–133). Charlotte, NC: Information Age Publishing.
- Mischel, W. (2004). Toward an integrative science of the person. *Annual Review of Psychology*, 55, 1–22.
- Mischel, W., & Shoda, Y. (1995). A cognitive–affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, 102, 246–268.
- Moors, A., & de Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, 132, 297–326.
- Nilsson, A. (2014). Personality psychology as the integrative study of traits and worldviews. *New Ideas in Psychology*, 32, 18–32.
- Panksepp, J. (2005). Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition*, 14, 30–80.
- Pashler, H., & Wagenmakers, E. J. (2012). Editors' Introduction to the Special Section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, 7, 528–530.
- Passmore, J. A. (1935). The nature of intelligence. *Australasian Journal of Psychology & Philosophy*, 13, 279–289.
- Petocz, A., & Newbery, G. (2010). On conceptual analysis as the primary qualitative approach to statistics education research in psychology. *Statistics Education Research Journal*, 9, 123–145.
- Salgado, J. F. (2003). Predicting job performance using FFM and non-FFM personality measures. *Journal of Occupational and Organizational Psychology*, 76, 323–346.
- Sijtsma, K. (2012). Psychological measurement between physics and statistics. *Theory & Psychology*, 22, 786–809.
- Skinner, B. F. (1953). *Science and behaviour*. New York: The Free Press.
- Smillie, L. D., Pickering, A. D., & Jackson, C. J. (2006). The new reinforcement sensitivity theory: Implications for personality measurement. *Personality and Social Psychology Review*, 10, 320–335.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103, 677–680.
- Terracciano, A., & McCrae, R. R. (2012). Why do (some) birds flock? Causality and the structure of characteristic adaptations. *European Journal of Personality*, 26, 449–450.
- Thomson, G. H. (1951). *The factorial analysis of human ability*. London: University of London Press (originally published 1939).
- Vautier, S. (2011). The operationalization of general hypotheses versus the discovery of empirical laws in Psychology. *Philosophia Scientiae*, 15, 105–122.
- Vautier, S., Lacot, E., & Veldhuis, M. (2014). Puzzle-solving in psychology: The neo-Galtonian vs. nomothetic research focuses. *New Ideas in Psychology*, 33, 46–53.
- Vautier, S., Veldhuis, M., Lacot, É., & Matton, N. (2012). The ambiguous utility of psychometrics for the interpretative foundation of socially relevant avatars. *Theory & Psychology*, 22, 810–822.
- Wakefield, J. C. (2007). Why psychology needs conceptual analysts: Wachtel's 'discontents' revisited. *Applied & Preventative Psychology*, 12, 39–43.
- Yankeeovich, D. (1973). The idea of human nature. *Social Research*, 40, 407–428.