Erkenn (2013) 78:739–761
DOI 10.1007/s10670-013-9502-z

ORIGINAL ARTICLE

Psychological Measurement and Methodological Realism

S. Brian Hood

Received: 13 May 2011 / Accepted: 13 June 2013 / Published online: 22 June 2013

Abstract Within the context of psychological measurement, realist commitments pervade methodology. Further, there are instances where particular scientific practices and decisions are explicable most plausibly against a background assumption of epistemic realism. That psychometrics is a realist enterprise provides a possible toehold for Stephen Jay Gould's objections to psychometrics in *The Mismeasure of Man* and Joel Michell's charges that psychometrics is a "pathological science." These objections do not withstand scrutiny. There are no fewer than three activities in ongoing psychometric research which presuppose a commitment to a minimal epistemic realism. Those activities include selecting between different models for representing data, estimating ability in the context of item response theory, and the move to make the individual the fundamental unit of analysis in psychometrics thereby calling for a shift in what sorts of data are evidentially relevant. In none of these activities are the commitments and disregard for evidence that Gould and Michell find objectionable or "pathological."

1 Facets of Realism

Statements of scientific realism are diverse, though they tend to claim something along the lines of the following: the central claims of our most successful scientific theories are at least approximately true. Unpacking this thesis typically involves laying out scientific realism's various components including a semantic thesis (semantic realism), a metaphysical thesis (metaphysical realism), and an epistemic thesis (epistemic realism). Also characteristic among statements of scientific realism are claims about the "aims" of science, e.g., science aims to provide true descriptions of the world.

S. B. Hood (✉)
University of West Florida, 11000 University Parkway, Pensacola, FL 32514, USA
e-mail: shood@uwf.edu

Semantic realism asserts that scientific claims about unobservable entities are to be taken literally and those claims are either true or false. Furthermore, the truth of an empirical claim is distinct from the evidence one has for asserting that claim. As Psillos (1999, 12) states, "'For every quark there is a corresponding lepton' is true if and only if for every quark there is a corresponding lepton." Semantic antirealists would deny this claim, perhaps asserting that 'For every quark there is a corresponding lepton' is true if and only if C, where C would be some statement of immediate sense experience. Other semantic antirealists may deny that scientific claims about unobservable entities admit of determinate truth values. I will not say much more about semantic realism except that it is a common ground among some scientific realists and antirealists.

Metaphysical realism asserts the existence of a mind-independent world. According to this view, the way the world is is independent of the theories scientists use to describe it. Moreover, metaphysical realism denies that the world consists of mental entities only (idealism) or that there is a distinction between the world as we investigate it and the way it *really* is. A detailed discussion regarding the controversies concerning semantic and metaphysical realism is beyond the scope of this paper. The focus here is epistemic realism and particular ontological commitments in psychometrics.

Epistemic realism, the thesis that we are epistemically justified in believing the claims of our most successful scientific theories, is where many contemporary scientific realists and antirealists find themselves divided. "Success" is typically understood to refer to predictive success. There are two main lines of argument that threaten epistemic realism. One line of argument concerns the under determination of theory by data: for any data that are predicted by some theory, T, there is some empirically equivalent rival theory T* which also predicts those data. Since we are not in a position to adjudicate between T and T* on empirical grounds alone, T's predictive success is insufficient grounds for believing T over T*. This is where pluralism gets its foothold as well as those would advocate the introduction of nonepistemic values into theoretical disputes (for example, Longino 1990, 2002). Another line of argument is grounded in the history of predictively successful theories that have nevertheless turned out to be false. Advocates of this argumentative strategy reason that predictive success is no indicator of truth since the history of science readily offers examples of theories whose central theoretical posits do not exist despite the success of the theories that posit them. In response to these challenges realists have considered numerous strategies, one of which has been to retreat to a weaker form of epistemic realism which I will call, following Leplin (1986, 1997), "minimal epistemic realism."[1] Minimal epistemic realism (henceforth, MER) can be characterized as follows:

> *MER*: it is, in principle, possible to justify hypotheses that posit unobservable entities.

[1] While Leplin's arguments for MER engage the historical argument, he argues against the underdetermination thesis on the grounds that the principle argument for it is incoherent (see Leplin 1997).

Note that MER is consistent with agnosticism about any particular theoretical posit. This is an important point. Some methodological decisions require a commitment only to MER, stopping short of ontological commitment. The focus of this paper is on methodology as implicated in theorizing to emphasize the role of realism in the *development* of theories and not theory adjudication itself. The latter is the object of scrutiny for those who claim that under determination undermines epistemic realism, i.e., those who argue that theory adjudication cannot be grounded solely in empirical data, and that, therefore, theory choice must appeal to non-epistemic virtues.[2] The truth of MER is not under dispute here; what is under dispute is whether MER is a necessary presupposition for the rationality of some scientific decisions.[3]

The terms 'rationality' and 'rational' are loaded and beg to be unpacked. Rationality, in so far as I will be employing the concept, requires that the processes, including the decisions, operating in the course of scientific theorizing be explicable in terms of justifiable general principles. That the principles be justifiable means that they can grounded either in terms of achieving some theoretical goal (such as truth, empirical adequacy, or theory articulation) accepted by the relevant scientific community or in virtue of their conformity with the canons of legitimate inferential practice as set forth by the relevant scientific community. I am not concerned with uses of 'rational' as applied to individuals, e.g., "Sam is rational," including assessments of individual's particular decisions, e.g., "Sam's decision to use experimental design such-and-such is rational." As I am using the term, the rationality of Sam's decision to use experimental design such-and-such depends on whether that decision can be explained in terms of its serviceability to some theoretical goal or in terms of its conformity to the standards of good inferential practices set forth by the scientific community to which Sam belongs. If the lessons drawn from the case studies below are correct, then a presumption of MER is required to explain certain methodological moves in psychological research. I will be offering, therefore, an indispensability argument not for MER but for the presumption of MER in the methodology of psychological science. That is, I argue for *methodological realism* defined thus:

> *Methodological realism*: the rationality of certain methodological decisions requires a commitment to MER.[4]

Methodological realism and MER are logically independent; it may be that MER is false, but that a (misplaced) commitment to MER is required to render methodology rational. MER is also logically independent of ontological realism with respect to some entity or class of entities. I define what it is to be an *ontological realist* about some entity, or type of entity, thus:

[2] Some, including McMullin (1983, 1987), have argued that such supplementary criteria may be considered epistemic virtues nonetheless.

[3] For a defense of MER, see Leplin (1997).

[4] This term is Leplin's (1986). He argues for methodological realism in Millikan's determination of elementary charges, relativistic cosmology, and quantum chromodynamics.

Ontological realism (about some specific entity or class of entities): commitment to the existence of some specified entity or entities of a specified type.

Ontological realism concerns specific ontological commitment. Of course, here the relevant entities will be those at issue in realism debates, namely unobservable entities, and in the context of psychological measurement the relevant entities are psychological attributes. Since methodological realism may be true even if ontological realism is false and since ontological realism may be false even if methodological realism is true, they too are logically independent; therefore, methodological realism does not require attributing presumptions of ontological realism. Likewise, MER and ontological realism are logically distinct; MER is an epistemological thesis, while ontological realism is a metaphysical thesis. One may be epistemically justified in believing that an entity exists, but that belief may nevertheless be false. Thus, the rationality of certain methodological decisions may coherently presuppose a commitment to the *possibility* of evidence accruing in favor of hypotheses that posit unobservable entities (MER), and yet those entities may not exist. Moreover, there need not be any tension between agnosticism regarding a particular posit, or class thereof, and MER; such selective agnosticism is consistent with commitment to the possibility that evidence could make one a believer (or nonbeliever). This latter commitment is tantamount to renouncing dogmatic agnosticism.

In what follows, I argue that that psychometrics, as practiced, is unintelligible if we do not ascribe realist commitments to psychometricians. For example, there are practices in psychometrics, including the process of theorizing, which require a commitment to MER, thus vindicating methodological realism. Indeed, the intelligibility of some psychometric practices arguably requires a commitment to ontological realism. Therefore, my arguments are directed, first, at those who would argue that there are rational reconstructions of psychometric practice purged of realist commitments; second, at those practicing psychometricians who claim that latent variables are merely mathematical transformations and do not denote, nor should they be taken as possibly denoting, psychological attributes; and, third, at critics of psychometrics such as Gould and Michell who claim that psychometricians are metaphysically committed in ways that are not scientifically legitimate or, in Michell's words, "pathological." Epistemic antirealism, its tenability notwithstanding, cannot be squared with the rationality of psychometric practice. If methodological realism is true, then the fact that psychometricians typically proclaim agnosticism or antirealism obscures their epistemic aspirations. However, the commitment to realism which I argue is necessary in certain psychometric practices is not of the sort that Gould and Michell find offensive.

To aid in framing the discussion, I first provide a brief introduction to psychometrics, in particular the assessment of cognitive ability. I then state the charges of Gould and Michell in terms of the various realist theses. In arguing for methodological realism, I consider three case studies directly implicated in the psychometric study of cognitive ability. The first case concerns model selection in representing the structure of ability using interindividual (i.e., population-level)

models of variation. The second case concerns two activities, prediction and estimation, as conducted in the context of item response theory (IRT), specifically, predicting performance and estimating ability. The third case illustrates that a presumption of MER is required for understanding what motivates moving from interindividual models of variation to intra individual (i.e., person-specific longitudinal) models of variation. This latter case, while similar to the first in its consideration of model selection, represents a debate over standards of evidence.

2 Psychometrics

Psychometrics, the branch of psychology devoted to psychological and educational assessment, is usually considered to be coextensive with 'intelligence testing', though the purview of psychometrics is much wider and includes personality assessment and the development of statistical models for research in the social sciences. At its inception in the early twentieth century the field was unabashedly realist. Constructs, the provenance of which was statistical analysis, were hypothesized as referring to psychological attributes, i.e., dispositions to manifest some behavior. For example, Spearman and others observed that when a diverse battery of ability tests were administered to a group of individuals, the correlations between all the tests were positive; those who did better on one type of test generally tended to do better on other types of tests even when the tests manifested no apparent overlap in content. They interpreted the positive correlations between different tests of intelligence (i.e., the "positive manifold") as indicating, if not referring to, an unobserved (or "latent") common cause of the positive manifold. This purported common cause manifested itself statistically as a general factor that accounted for (i.e., screened off) an appreciable portion of the correlations between measures of mental ability such that if one controlled for variation in the general factor, the correlations between the different measures disappeared. Spearman hypothesized that this general factor, or *g*, referred to a property of the mind, namely "general intelligence," (1904, 1927). In terms of the distinctions made in the previous section, Spearman was an ontological realist regarding general intelligence. He was also an epistemic realist, for he believed that statistical techniques such as factor analysis provided evidence for the existence of general intelligence and described its structure. With the ascent of logical positivism, psychometrics assumed a different philosophical orientation. Operationalism, a form of semantic antirealism, took hold. In psychometrics, if not in psychology more broadly, constructs were defined in terms of measurement instruments or in terms of their statistical relationships with other tests (Cronbach and Meehl 1955).[5] Hence, in the

[5] For example, Boring, when asked to define 'intelligence', famously claimed "Intelligence is what intelligence tests test," (Boring 1923). Even psychometric concepts were reconceptualize in accordance with the extreme empiricism that pervaded psychology. The concept of psychometric validity, once cashed out in terms of whether a test measures the attribute it purports to measure (Kelley 1927), later was formulated in terms of whether performance correlated with some other criterion such as academic success, and likewise for psychological constructs: they too were defined purely in terms of their statistical relationships with observable criteria.

service of empiricism's ontological austerity, attributes such as general intelligence were stripped of their psychological content by being defined purely in terms of their syntactic relationships with other quantities. Under this scheme, general intelligence would be implicitly defined as that thing which correlates such and such with occupational success or academic achievement, not "mental energy" as Spearman conjectured.[6] Empiricists of this tradition strictly delimited their credence to that which was observable, thereby denying epistemic realism.

In the latter half of the twentieth century psychometrics continued to flourish. Psychometricians invented new statistical tools for fine-grained data analysis. Item response theory (IRT) shifted the focus of analysis from the test as a whole and batteries of tests to individual items, and its stochastic elements provided refined techniques for analyzing measurement error. IRT also suggested new ways of thinking about the functional relationship between item parameters (such as difficulty) and levels of the psychological attribute purportedly being measured. IRT introduced novel ways of estimating ability and predicting performance. More recently, psychometricians have sought to shift the unit of psychometric study from the population to the individual. Traditional psychometric methods involved administering tests to large groups of individuals, noting correlations, and positing some common source for the score variability. The associated statistical models represent variability *between* individuals; however, with rising concerns over what such models entail for the *individual* and individual development, psychometricians have constructed models that analyze the individual over time. The resulting models, intraindividual models, are aimed as discovering processes at work within the individual. After all the structure of variability between individuals, the structure that grounds the theory of general intelligence, may not fit any particular individual in the population.

There seems to be a shift back to a more "realist" psychometrics. There are obvious examples of this turn (see Borsboom 2005; Borsboom et al. 2003), but I wish to argue for something more significant than that there are some psychometricians who openly endorse realism in some form; specifically, in psychometrics the rationality of methodological decisions and practices implicated in theory formation is undermined unless we attribute at least MER to psychometricians, regardless of the philosophical positions that particular psychometricians openly espouse.

3 Psychometrics as "Pathological Science": Gould and Michell

In *The Mismeasure of Man* Gould argues that since the development of factor analysis and Spearman's "discovery" of *g* in 1904, psychometricians unreflectively

[6] Schlick's method of implicit definition (Schlick 1918) finds its inspiration in Hilbert's work in the foundations of geometry (Hilbert 1899) and Einstein's and Poincaré's work in theoretical physics and the philosophy of geometry (Feigl 1970). See Friedman (1999, especially 34–35) and Coffa (1991, especially 176–177) for an insightful discussion of the role of implicit definition in Schlick's epistemology. For Schlick, it is by implicitly defining central theoretical concepts, and hence divorcing them from immediate sense experience, that objective knowledge is possible. The method of implicit definition also finds a home in Carnap's *Aufbau* (1928). Later logical empiricists such as Feigl (1970, 7) defend the method of implicit definition for theoretical concepts the meanings of which "can be specified only by their place in the entire theoretical system...[.]"

and blatantly reified *g* as general intelligence. Psychometricians have responded that Gould's knowledge of psychometrics is decades out of date, and by professing agnosticism about the nature and existence of attributes, regarding them as mere hypotheses. Psychometricians are correct to reject Gould's accusations that they naively reify latent variables (as psychological attributes). They counter Gould by citing their own lack of ontological commitments and their epistemic goals (see Carroll 1995; Deary 2000; Nyborg 2003; Bartholomew 2004).[7]

If central practices and methodological decisions in psychometrics such as estimating ability, model selection, and shifting standards of evidence can be rationalized only if realism is presupposed, then there might yet be grounds for Gouldians to object. In Sect. 4, I argue that methodological realism does find support in these practices. Consequently, the Gouldian argument against psychometrics may be reformulated as a critique of the underlying epistemic commitments guiding practice. That is, using methodological realism as a toehold, the Gouldian may find a position from which to attack MER (with respect to psychometric theoretical entities). The argument would proceed along the following line: since methodological realism is true for the relevant psychometric practices (such as measuring cognitive ability), the rationality of these practices assumes the truth of MER; however, MER is false. Thus, psychometrics is presented with a dilemma: continue relying on a false presumption of MER, or reject MER and render many of its practices seemingly arbitrary. This argument, proposed on Gould's behalf, is a much more serious objection since it implicates specific modern psychometric practice, not simply a handful of its practitioners. Thus I have sought to update Gould's objection in response to those who would complain that the target of his critique is a straw man. I will assess this Gould-inspired objection and argue that methodological realism in the context of psychometrics is unobjectionable.

A brief note regarding Gouldians and methodological realism: they are not likely to balk at the result that methodological realism finds support in much of psychometric practice. The philosophical significance of this result is that psychometrics cannot be rationally reconstructed without realist commitments, and Gould would likely agree with this assessment of psychometrics even if he does not himself endorse any of the psychometrician's substantive theoretical claims. Were I to *defend* epistemic or ontological realism in psychometrics, Gould and I would have to part ways since his attack on the reification of latent factors aims to undermine realism, both ontological and epistemic. The spirit, if not the letter, of Gould's objections seems correct: the unsubstantiated reification of mathematical entities seems at odds with norms governing legitimate scientific practice. As methodological realism in the context of psychometrics holds that realist presuppositions underpin psychometric research, it seems that Gould's criticisms may yet have a target, psychometricians' professed epistemic and ontological modesty notwithstanding.

Michell (1997, 1999, 2000, 2004, 2008) argues along similar lines, claiming that psychometrics is an instance of "pathological science." The basis for this critical

[7] Elsewhere I have argued that it is doubtful that Gould's charges hold even for Spearman and Jensen, two psychometricians to whom Gould devotes much effort to portray as pathological "reifiers."

diagnosis is that psychometricians, without evidence, assume that psychological attributes have a quantitative structure. That is, in terms of levels of measurement, psychological attributes can be measured on interval scales (e.g., temperature measured in degrees Celsius) or ratio scales (e.g., length) such that there a functional relationship between quantitative characteristics of observed scores and quantitative characteristics of the attribute. This is an empirical assumption since differences in test scores may be the result of factors that are not quantitative in character, for example the content of one's prior educational experiences. The pathological aspect of making such an assumption is that, first, the assumption is never tested and, second, that mainstream psychometricians ignore the fact that the assumption is never tested (Michell 2000, 2004, 2008). The second symptom of pathology is tantamount to an accusation that psychometricians are unresponsive to criticism. Michell's critique, thus, seems to presuppose that those who are doing pathological science, at least in this case, harbor (unwarranted) realist commitments regarding the properties of psychological attributes. Michell is describing a situation in which psychometricians ascribe properties to those things about which they are ontological realists, though they do so without fulfilling the epistemic obligations that license such an ascription. In other words, Michell is an ardent proponent of methodological realism, attacking mainstream psychometricians for running afoul of methodological norms associated with ontological realism regarding psychological attributes. Because the psychometricians are ontologically committed to psychological attributes and they ascribe certain properties to them, namely being quantitative in structure, Michell faults them for eschewing what he sees as a normative requirement to gain theoretical knowledge about the theoretical entities in question rather than simply presume it. For Michell, with ontological commitment comes epistemic obligation. Ontological commitment in absence of MER would be sufficient for unresponsiveness to criticism. Without MER ontological commitments are unjustified and insulated from criticism; the claim that such theoretical commitments can be justified just is MER. If one does not accept that such claims can be justified, but nevertheless remains a realist about the relevant theoretical entities, evidence is rendered irrelevant in principle.

However, ontological realism in conjunction with MER seems to be *prima facie* unobjectionable, for it entails a willingness to allow empirical commitments to answer to evidence. There may be specific instances where science motivated by realism goes awry, but it is not obvious that it would be a general feature of the marriage of ontological realism and MER, rather than features peculiar to the example, that would be to blame. MER alone also seems entirely benign, for at worst, i.e., if epistemic antirealism is true, MER would be tantamount to a naïvely optimistic epistemic aspiration, but it would still place empirical evidence as a key arbiter in empirical disputes. MER may very well be false, and some may argue that if it is false, then we should not believe it. However, the sensible advice that we ought to avoid believing falsehoods is not the same as recommending that false claims should also be rejected as guiding methodological principles, for they may be justified as such on other grounds. For example, as a methodological principle guiding research, MER may be pragmatically justified in terms of its ability to further certain theoretical goals. MER may be a lie, but it may also be a useful one.

Further, if MER is unscientific or pathological, then much of what is regarded as exemplary science is rendered unscientific, and that is unacceptable.[8]

There is some controversy over Michell's claim that the quantity hypothesis is never tested. Borsboom and Mellenbergh (2004) argue, for example, that the hypothesis is tested every time a quantitative model of the relationship between an attribute and a score is tested, for the quantity hypothesis is an auxiliary assumption. The motivation for this response is the Duhem-Quine thesis according to which no hypothesis is testable in isolation and that confirmation is a holistic affair such that confirmation of the model entails confirmation of the auxiliary assumptions which, in conjunction with the model, entail a set of predictions.[9]

For those epistemic (or semantic) antirealists who doubt that certain methodological practices require (or are at least most plausibly interpreted as motivated by) realist presuppositions, note that antirealism provides no safe haven. Antirealists may respond to Michell's critique claiming either that they do not believe that quantitative psychological attributes exist or that we cannot justify claims about such entities; they will seek exemption, claiming that they are uncommitted to the existence of quantitative attributes. However, Michell's critique stands, if only at the level of methodology. The target of the critique can be shifted from the psychometrician's ontological commitments to methodological practice in order to undermine the antirealist response. In Sect. 4, for example, I argue that the practice of estimating cognitive ability presumes realism. Suppose this is false, however, and, the estimation of ability can be plausibly interpreted in an antirealist light. The antirealist response would go something like this: the practice of "estimating ability" does not indicate a commitment to the existence of the latent ability; what it indicates is a preference to model data *as if* there were such an ability. Many psychometricians would protest that this grossly distorts how they think of measurement models. Nevertheless, it is conceivable that a philosopher occupies such a position. Michell could so tailor his objection: proceeding *as if* the traits were quantitative is no better than assuming that they are. From a methodological standpoint there is no difference. The quantity hypothesis does the same work in the models even if it does not require a realist interpretation; the hypothesis that traits are quantitative, too, requires testing just as any other auxiliary hypothesis would. But going along with the "*as if*" antirealist even this far might be too charitable. It is unclear how the "*as if*" antirealist could make sense of testing the quantity hypothesis without being committed to at least MER. Regardless, if "*as if*" antirealism is a coherent position in this context, Michell's critique stands.

Gould and Michell's general and very sensible idea that unresponsiveness to criticism represents a breakdown of scientific objectivity finds considerable support in recent work in the philosophy of science (see Fagan 2007; Kitcher 2001; Lloyd 2005; Longino 1990, 2002). Their critique requires that we take a closer look at the role realism plays in psychometric research. The next section considers three cases

[8] Hood (2008) challenges the very notion of "pathological science" arguing that it is not only vague, but that it is unclear what the appropriate course of action is for a scientist engaged in pathological science.

[9] This response stands or falls with holistic confirmation, and a critical evaluation of this thesis is beyond the scope of this paper.

in which realist presuppositions are operating. In each case I distinguish whether the relevant assumption is epistemic, ontological, or both. The assessment of these cases as fundamentally realist is a philosophically significant result in its own right, apart from its relevance to Gould and Michell's critique.

4 Methodological Realism in Psychometrics

The rationality of certain psychometric methodological decisions requires a commitment to MER, and in some cases a further commitment to ontological realism regarding psychological attributes is required. The justification for these commitments will not be considered here. Rather, I argue that these commitments, either to minimal epistemic realism or to ontological realism, render comprehensible, if not explain uniquely the following activities in psychometrics: model selection, estimation of ability, and choice of experimental design as it specifies standards of evidence.

4.1 Model Selection and the Representation of Psychological Attributes

Borsboom (2005) articulates a compelling piece of evidence in favor of methodological realism, arguing that the preference for certain kinds of measurement models in psychology suggests underlying realist commitments among psychologists and psychometricians. Consider the two different psychometric models for representing the structure of individual differences in Fig. 1.

Following Edwards and Bagozzi (2000) and Borsboom (2005), I will refer to the model on the left side of Fig. 1 as a *reflective* model and to the model on the right as a *formative* model. Each $\mathbf{X_i}$ is a manifest (i.e., observed) variable or indicator such as an item response or test variable. ξ and η are latent variables, each λ_i is the factor loading of each indicator in the left-hand model, and each γ_i is a weight of the

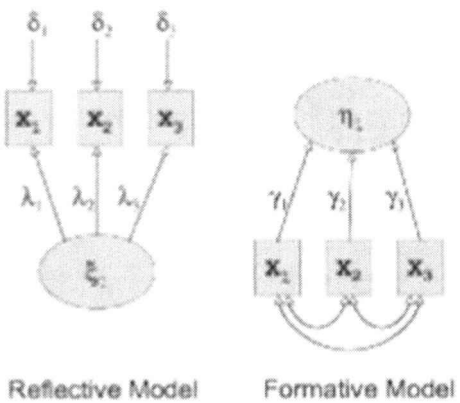

Fig. 1 Reflective and formative measurement models

indicator with respect to the latent variable.[10] Each δ_i is an error term for the relevant indicator.

The reflective model is the typical unidimensional measurement model found in psychometrics; it posits one unobservable dimension, variation along which purportedly causally explains variation along the various observable measures. In the measurement of general intelligence, each $\mathbf{X_i}$ would be, for example, a subtest (or item) of the Wechsler Intelligence Scale for Children (WISC-IV) and performance on each subtest (or item) would be seen as a function of position on the latent variable *g*; it is differences in positions on *g* which cause differences in performance on the indicators, hence the direction of the arrows. For example, suppose we have two individuals, Nino and Eka, who take the WISC-IV. The psychologist performing the assessment would likely attribute the differences in scores to differences in general intelligence, represented by a *g*-score. Formative models, on the other hand, are popular in sociological research. For example, socioeconomic status (SES) is often modeled formatively. In the formative model the direction of causal influence is reversed, running from the indicators to the latent variable. The latent variable is regressed on its indicators, not the other way around. One (or a population) occupies a position on SES because of the indicator values, such as occupational prestige, and SES is interpreted as summary of those values; no ontological commitment is required. We may even use one's SES score to predict one's level on some unmeasured indicator, but even this does not entail that SES is being treated as existing independently of its indicators. Representing individual differences on the WISC-IV with a formative model reverses the direction of causation posited in the reflective model. If we interpret the latent variable as general intelligence (which would just be some weighted composite of the test scores) and interpret the arrows causally, then it is differences in performance on the tests that cause Nino and Eka to have different *g*-scores. In other words, in reflective models, individual differences in general intelligence causes differences in test scores; in formative models, differences in tests scores cause differences in general intelligence. It is worth noting that psychometricians never model cognitive ability formatively, and to identify the *g* of a reflective model with the *g* of a formative model would be dubious since they would share no mathematical characteristics and would occupy opposing roles in causal explanations of individual differences.

The choice to model covariation with a reflective model, as opposed to a formative model, requires a commitment to realism. There is no a priori reason to prefer one model to the other unless one harbors realist commitments. Without such commitments, the decision seems arbitrary. Indeed, it would seem irrational to prefer the model that carries with it ontological commitment if one does not embrace some form of realism. Borsboom claims that model selection requires both ontological realism and MER in order to be non-arbitrary. That is, he believes that reflective model selection requires commitment to the existence of the attribute as well as the possibility of evidence accruing in favor of hypotheses that posit the attribute; however, it seems that the demands of rationality require only the latter.

[10] The factor loading of indicator X with respect to a factor F is an estimated (Pearson) correlation between X and F.

Psychometricians may commit themselves to psychological attributes being the referents of latent variables; however, choosing to model covariation reflectively does not seem to *require* them to do so. Measurement models, such as those in Fig. 1, are *hypotheses* and as such they may enjoy more or less empirical support. They are offered up as explanations of variations in test performance in a population. Proposing a hypothesis does not *ipso facto* commit one to the existence of the entities postulated therein. That psychometricians model covariation reflectively while remaining agnostic about the ontological status of their latent variables is perfectly intelligible; one might even argue that it is just good science to withhold ontological commitment until sufficient evidence has accumulated that would warrant belief in latent traits. Psychologists may nevertheless attempt to justify their ontological realism by appealing to the model's ability to fit the data, or appeal to the explanatory or predictive power of a particular model, but there is nothing in the process of modeling or selecting a reflective model that carries ontological commitment. While we can make sense of model selection without attributing ontological realism to psychometricians, the rationality of model selection seems to require attributing MER, hence confirming methodological realism.

Reflective models are hypotheses about the structure of covariation. They are also causal hypotheses about the underlying, i.e., latent, common cause(s) of test behavior. These models are selected not only for their structural and mathematical properties, but because they are testable against novel data through confirmatory factor analysis. A fit between the model and data is interpreted as confirmation of the model *qua causal model*. A good example of such a study is Jensen and Weng's (1994) study where the authors sought to verify the presence of a dominant latent factor, viz., *g*, in a variety of samples, both simulated and real, through exploratory factor analysis and confirmatory factor analysis. It would seem to make little sense to be "testing" these measurement models against data if Jensen did not think that there was the possibility of confirming the *g* hypothesis. Were his sole aim empirical adequacy and not ascertaining the structure of mental ability, the selection of reflective measurement models would make little sense, as there are available formative models that also capture the data. Merely subjecting such models to a confirmatory factor analysis suggests that one is aiming to confirm the model, causal relationships, unobservables, and all. Thus, the selection of models that posit latent sources of variation and the subsequent testing of those models, as opposed to models that do not make such empirical claims about the latent structure of ability give good grounds for methodological realism. Psychometric practice is arbitrary unless we attribute to psychometricians (at least those who work on *g*) the belief that evidence can accrue favorably or unfavorably for the existence of psychological attributes.

That Jensen and Weng appeal to simulated data may seem to be problematic for methodological realism. Simulated data sets and the analysis thereof would not, on the fact of it, require MER since the object of study is, by design, a statistical artifact with no connection to psychological phenomena. Had Jensen and Weng used only simulated datasets, admittedly MER would not be required in this case. However, Jensen and Weng's motivation for using simulated datasets in conjunction with data gathered from subjects betrays a commitment to epistemic realism. Their goal is to demonstrate the robustness of *g* across different methods of factor analysis. They do

this by applying six different methods of factor analysis to four contrived correlation matrices whose factor structure is already known. Several of the simulated matrices are designed to mislead some of the six methods into reporting *g* loadings that deviate from the true *g* loading; however, the resulting deviations from the true factor loadings turn out to be negligible. The average loading on *g* is .5 on the simulated matrices and the deviations range from .031 to .059. Hence, *g* appears robustly and reliably across diverse methods in the case of the simulated data. Additionally, ten methods of factor analysis are applied to a correlation matrix derived from actual scores on tests, specifically the performance of 145 seventh and eighth graders on 24 diverse mental ability tests. Of course, there is no value for the true loading on *g* as in the simulate matrix. Nevertheless, Jensen and Weng find that the estimates of *g* loadings across the ten different methods have correlations that ranged from .991 to 1.00. Based on the result of the simulated matrix study, i.e., that strong agreement is present when the tests all closely estimated the true *g* loading, Jensen and Weng infer, with respect to the 24 tests, that the reported estimates of *g* loadings also closely estimate the true *g* loadings for the matrix. Thus, the role of the simulated dataset and its analysis is to lend credibility to the hypothesis invoked to explain the results of the subsequent analysis of non-simulated data. Were they not at least committed to MER, it would be difficult to make sense of the appeal to idealized data. It would also be difficult to make sense of such claims as the following where they admit that

> factor analysis would have little value in research on the nature of abilities if the discovered factors did not correspond to real elements or processes, that is, unless they had some reality beyond the mathematical manipulations that derived them from a correlation matrix (Jensen and Weng 1994, 254).

This passage, and indeed the entire study, would seem rather incomprehensible if Jensen and Weng do not believe that they are making discoveries about the nature of *g*.

4.2 Item Response Theory: Prediction and Estimation

A related but distinct example of the role of methodological realism in psychometrics concerns item response theory (IRT). 'IRT' refers to a class of models used to predict test behavior on the basis of position on some latent variable and properties of particular test items. Such properties include item difficulty, the item's capacity to discriminate between individuals of different ability, and the probability of guessing the solution to the item. IRT models also provide a means for estimating ability. IRT was developed in response to many problems that plagued classical test theory. An IRT model specifies an item characteristic curve (ICC), a monotonically increasing function that describes the relationship between position on a latent variable and test performance (Hambleton et al. 1991). That an ICC can describe performance is one postulate of IRT. The other is that performance can be predicted by position on the latent variable. There are further relevant assumptions of IRT models. First among them is *unidimensionality*. A set of items is unidimensional just in case the item(s) in that set measure only one

attribute. In many IRT texts, including Hambleton et al.'s canonical introduction to the subject, unidimensionality is treated as an assumption of IRT models.[11]

An important *desideratum* for IRT models is that they exhibit *local independence*. A model is locally independent when the correlation between performance on test items disappears when conditionalized on the latent variable—that is, the latent variable *screens off* the correlations between item responses much like the falling of a barometric pressure screens off the correlation between rain and falling barometers. Formally, we may express local independence with the following equation:

$$\Pr(R_1, R_2, \ldots, R_n|\theta) = \Pr(R_1|\theta)\Pr(R_2|\theta)\ldots\Pr(R_n|\theta) = \prod_{i=n}^{n}\Pr(R_i|\theta)$$

where 'Pr' denotes the probability function, 'R_k' denotes a response on some item, and 'θ', a latent variable, denotes an unobservable psychological attribute. This equation states that the probability of a set of responses given the latent variable is identical to the product of every response conditional on the latent variable. Hence, there is a close conceptual connection between an IRT model and the reflective model. The latent variable screens off the correlations between subtests and items in the reflective model and IRT model, respectively. Additionally, variations in values on the latent variable are generally treated as being conceptually prior to variations on items or subtests in the sense that the former are hypothesized to cause the latter. IRT models are simply probabilistic reflective models of variability in individual items (instead of subtests, which are aggregates of items).

The following equation is an example of a one-parameter logistic IRT model:

$$\Pr_i(\theta) = e^{(\theta - b_i)}/1 + e^{(\theta - b_i)} i = 1, 2, 3, \ldots, n$$

where $\Pr_i(\theta)$, an S-shaped curve with value between 0 and 1, is the probability of answering item i correct given latent ability θ, b_i is the difficulty parameter of item i, n is the number of items, and e is a constant (2.718). In this case where $n = 4$, this model defines the following ICC for four items, which is reprinted from Hambleton et al. (1991, 14).

Additional parameters could be added, such as a discrimination parameter, or a guessing parameter (also called "pseudo-chance"), depending on the data to which one is fitting the model; however, for the present purposes (thinking about the status of θ), the above example is sufficient. Figure 2 demonstrates how the probability of correctly responding to four distinct items (the vertical axis) is a function of the level of θ (the horizontal axis).

[11] It is arguable, however, whether unidimensionality is appropriately called "an assumption." Models are testable with respect to unidimensionality. If the posited latent trait accounts for test performance, the assumption is satisfied. If it does not, then the data may need more than one latent trait, and the resulting model is multidimensional. That the "assumption" is directly tested in fitting models, suggests it is not properly called "an assumption." Nothing in my argument hinges on the status of unidimensionality in this regard.

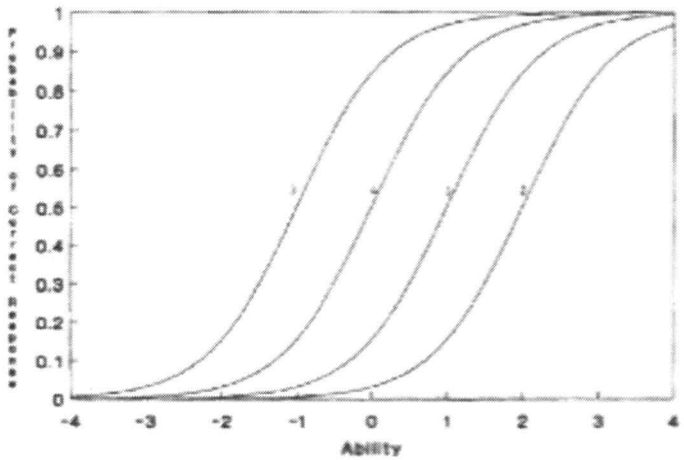

Fig. 2 Item characteristic curves for four items modeled by a one-parameter logistic IRT model

The question is this: is there a feature of IRT models and their place in psychometric theorizing that would commit psychometricians to ontological realism or MER? At first glance, the answer seems to be 'yes' on both counts. Ontological realism is required by the fact that performance is regressed on latent ability θ. MER is required by the fact that models are subject to testing, and, therefore, commitment to the possibility of evidence accruing in favor of models that posit latent traits. However, things are not so simple.

4.2.1 Predicting Performance Does Not Require Ontological Realism

Ontological realism with respect to psychological attributes, including general intelligence, is *not* required for the use of IRT models to be rational. As in the case of reflective models, IRT models may be considered hypotheses albeit ones that posit unobservable entities. However, conjecture need not carry commitment to the existence of the thing conjectured even if it does entail a commitment to the possibility that the conjectured posit exists. Accepting the model as providing a true account of the phenomena would require commitment to the existence of the psychological attributes only if 'θ' is read as denoting cognitive ability. If one reads 'θ' as a mere placeholder for confounding causal factors yet to be disentangled, commitment to psychological attributes may be eschewed. Even acceptance of the model does not seem to command commitment to the referential success of 'θ'. The acceptance of IRT models seems to require nothing more than commitment to the empirical adequacy of the model. There is no obstacle to treating IRT models as merely useful predictors of test behavior, though this suggestion will seem, if not patently false, strange to psychometricians and proponents of IRT, or anyone who has read an elementary textbook on IRT since it is typically listed among the assumptions of IRT models that there are underlying factors that explain test performance and these factors are the latent factors that appear in IRT models, i.e., θ (for example, see Hambleton et al. 1991). The suggestion here is that such an assumption is unnecessary. While I may

seem to be pitting myself against psychometric dicta, in fact I am exonerating psychometricians (or at least offering them an escape) from charges that their methods require the unreflective reification of latent factors.

4.2.2 The Estimation of Ability Requires Ontological and Epistemic Realism

There are certain practices in IRT modeling and latent variable theory that require at least a tacit commitment to epistemic realism if not full-blown ontological realism. Again, since I am investigating the presuppositions of psychometrics and not their justification, realism's tenability will not be addressed.

Given a function describing an ICC, a specification of parameters such as item difficulty, and a level on the latent variable, the psychometrician can predict performance on items. However, it would indeed be strange if tests were administered for the sole purpose of determining (probable) performance on other test items. The alleged purpose of ability tests is to measure ability. After all, what is the value of saying that John will probably get item 1 correct on test X given that he has ability level A as measured by test Y? This seems to get the testing enterprise backward. I don't deny that there is a legitimate place for such a practice in test development. For example, if the aim were to confirm a model, it would be valuable to see if the model correctly predicted performance. But at the end of the day, once an item analysis for the test has been conducted, the model is confirmed, and the appropriate validation studies have been conducted, the test is then implemented as an instrument for assessing ability in the population. Many assumptions go into assigning scores to individuals with respect to some latent factor (e.g., that scores are normally distributed in the population) but their legitimacy will not be discussed here. Such a discussion would bear more on the justification of realism than on the present concern. If we grant those assumptions (or turn a blind eye to them) and just focus on what it means to estimate an individual's position on a latent variable, we find that realist assumptions underlie practice.

There are several procedures for estimating someone's g-score, and none of them are trivial. *Maximum likelihood estimates* of θ introduce an operational latent variable θ', which acts as a proxy for θ. Simply put, θ' is an estimate of the value of θ that would render a pattern of item responses most likely. Calculating maximum likelihood estimates is complicated and not always feasible. One may also search for an operational latent variable that renders the factor model locally independent, e.g., the sum score; this proxy variable is called a *sufficient statistic* for the latent variable (Bartholomew 2004). The idea is that once you have identified some function of the $\mathbf{X}_i$, i.e., the manifest test variables, such that it screens off their correlations, you have captured all the information contained in g. The sufficient statistic can then be used to place examinees on an ordinal scale.[12] The particulars of the various approaches to estimating position on some latent variable need not concern us. It is what these approaches have in common that is relevant to the current discussion: they all carry with them a presumption of ontological realism and a commitment to MER.

[12] Ordinal scales give only relative measures. They have neither a non-arbitrary zero point, nor do they permit the interpretation of distances between magnitudes. Ratio scales are not deficient in these respects.

Suppose a psychometrician administers a test battery, the WISC-IV for example, to two examinees, Nino and Eka. Also, suppose there is a sufficient statistic θ' for g. Eka scores significantly higher than Nino, i.e., she has a higher value for θ' than Nino. The psychometrician infers from this fact that Eka occupies a higher position on g than Nino. The psychometrician then infers, based on their estimated relative positions on g, that Eka is more intelligent than Nino. This latter inference will be trivial if individual ability were defined as position on g, or if relative ability were defined as relative position on g. We need not concern ourselves with the legitimacy of the latter inference though it should be noted that estimating ability in this way requires local homogeneity, the claim that the population-level structure of individual differences is isomorphic to the structure of intraindividual differences.[13] A sufficient statistic is identified for *population-level* data. The sufficient statistic is then imported into assessment of individuals. Whether local homogeneity is justified is beyond the scope of this discussion. However, the dispute over the justification of the assumption of local homogeneity provides further evidence in favor of methodological realism.

There does not seem to be an antirealist interpretation of the practice of estimating individual ability (or the ability of groups) available. Of course, if the psychological attribute does not exist or if belief in its existence is unjustified, then any claims of the form 'S's estimated level of some latent trait θ is n' will be unjustified. But the *practice* of estimating latent ability seems to presuppose that *there is* some dimension on which individuals are located and that claims about position on the dimension are corrigible: the more tests we give, the more confident we are regarding be that S's "true" position on θ. Estimation presupposes that there is a determinate value on which we are trying to converge (Borsboom 2005). If psychometricians did not believe that the attribute in question existed, and if they did not further believe that claims about individual position on the attribute were corrigible, estimation procedures and attempts to reduce sampling error by giving a diverse battery of tests would be a perverse and arbitrary enterprise. Consequently, estimating ability requires both methodological realism and a commitment to the existence of latent traits.

Attending specifically to the procedure where a sufficient statistic is sought, one finds further support for methodological realism. Some statistics will be "more sufficient" than others. The reason why a particular statistic is chosen as the sufficient statistic is that it is alleged to contain more information about the latent trait than its alternatives, thus attributions made on its basis are deemed more justified than those made on the basis of an alternative statistic.

4.3 Standards of Evidence and the Unit of Psychological Research

If a model of covariation is locally homogeneous, then if that model fits a population of examinees, then it also fits each member of that population. For example, if either

[13] Interpreted causally, the local homogeneity assumption claims that the population-level measurement model represents not only measurement at the population level, but also measurement at the level of individuals. In other words, whatever causes differences in test scores at the interindividual level is also responsible for differences in test scores at the intra individual level; therefore, if a test measures general mental ability at the population level, it also measures the same trait in each member of the population and each subpopulation.

model in Fig. 1 is locally homogeneous, then it fits not only the data we obtain from the population, it also fits each individual examinee over repeated measures. At this level of description, local homogeneity is just a syntactic relation between models of covariation. However, the interpretation of local homogeneity is that whatever the test measures in the population is also measured in each test taker and subpopulation of test takers (Ellis & van den Wollenberg 1993, 422; Borsboom 2005), i.e., the covariation models are interpreted causally as measurement models applying both at the level of populations and at the level of individuals. In testing and test development, local homogeneity is often assumed. However, the legitimacy of this assumption has come under fire in the past few decades and at least as early as Baldwin 1946.[14] In spite of this controversy, intelligence and personality theorists persist in making the assumption despite the fact that a failure of local homogeneity imposes serious limitations on a conception of intelligence as that thing '*g*' denotes. I contend that realism motivates research on local homogeneity and subsequent developments responsive to problems associated with a lack of local homogeneity.

The force behind a demonstration that a measurement model is not locally homogeneous is straightforward when one considers its interpretation. If the population model is not locally homogeneous, the heterogeneity suggests that what the test (or battery of tests) measures in the population is not identical to what is measured in each individual. In the fields of personality and ability assessment, the alleged objects of measurement are latent traits such as extroversion and general intelligence. Therefore, to the extent that a test is taken to measure an attribute in the population (or individual), ontological realism about those attributes is assumed. Moreover, to the extent that a failure of local homogeneity is supposed to be evidence that what is measured in the population is (or may not be) not what is measured in the individual, epistemic realism is assumed.

In response to the problem posed by local homogeneity, psychometricians have developed and implemented statistical tools for the study of individuals over time (see Cervone 2004; Hamaker 2004; Molenaar 1985, 1999; Molenaar et al. 2003). Analysis of intraindividual variability over time is conducted through *time-series analysis*. Hamaker in particular suggests that psychology should shift its focus from interindividual variability, studied by way of standard factor analytic procedures, to intraindividual variability, studied by way of dynamic factor models that represent change in an individual over time. The result is a model, perhaps not unlike the reflective model, but the data are gathered from the same individual over time rather than a diverse population of individuals. The thought is that through time-series analysis and its integration with traditional methods, we will gain better insight into psychological processes. For example, Hamaker writes,

> If we want to obtain knowledge about intraindividual psychological processes and about the lawfulness that underlie them, we must employ a technique that allows us to make statements about the structure of variability within the individual, rather than about the distribution of variables in the population...If the individual is our unit of analysis, the results [of population-level data] do

[14] Baldwin writes, "There is no assurance, however, that the organization of personality variables within the individual is accurately described by the pattern derived from group studies" (1946, 152).

> not inform us about what is going on at the level of the individual, but about the laws operating at the level of the population (Hamaker 2004, 7).

The above quotation is typical of those who advocate a turn from traditional nomothetic studies in which the unit of analysis is the population and the relevant evidence is synchronic population-level data, to idiographic studies in which the unit of analysis is the individual and the relevant evidence is data gathered through diachronic time-series analysis. However one need not be an advocate of time-series analysis to acknowledge the statistical fact that some models are not locally homogeneous. Time-series analysis and hybrid approaches such as Hamaker's have yet to gain widespread acceptance, but it is unclear whether this is for sociological reasons such as the complexity of the techniques, or because the techniques are methodologically inappropriate. Nevertheless, there does seem to be a significant contingent within psychometrics advocating an idiographic turn in research and analysis. My claim is that the impetus for this turn is realist in character. There is an irony, however, in that this methodological shift also represents an antirealist attitude toward standard interindividual models such as those presented in Sect. 4.1. However, acknowledging this does not weaken the case for methodological realism in the case of selecting reflective models over formative models.

Consider the epistemology of the idiographic turn: Hamaker's quote above wears its epistemic realism on its sleeve, claiming that "if we want to obtain knowledge" of individuals' psychological processes, then we ought to employ techniques appropriate for studying intraindividual change. Nevertheless, we need not take Hamaker's (or any other psychometrician's) word for it. After all, the assumption of MER may be otiose. The idiographic turn has no apparent scientific motivation, either epistemic or pragmatic, and is best explained by a commitment to the possibility of gaining evidence about intraindividual trait structure and how psychological attributes develop over time. Without this commitment, the idiographic turn seems unmotivated. It is highly implausible that psychometricians and psychologists would campaign for using incredibly sophisticated and complex techniques unless they, perhaps even implicitly, endorsed the possibility of accumulating evidence concerning the nature of psychological attributes. Taking the turn cannot be rationalized by appeals to simplicity or to pragmatics, for the techniques in question are not part of the psychologist's or psychometrician's standard toolset, nor are they easily implemented. The relative novelty of these tools renders implausible the possibility that the turn admits of instrumentalist explanation, for the family of techniques under consideration do not have a history of utility that could pragmatically justify selecting these novel methods over more entrenched ones such as standard interindividual factor analytic techniques. An additional alternative explanation for the turn may be that it is motivated by socio-ideological (nonepistemic) values, but there is no discernible connection between advocating dynamic statistical models or the models themselves, and nonepistemic ends such as racism, economic gain, or political influence.

Empiricists may argue that some other commitment than to MER could motivate the idiographic turn. The obvious candidate is empirical adequacy. Under this alternative description of the methodological shift, there would be no reference to

psychological attributes or unobservable developmental processes. Time-series analysis would be preferred for its superior ability to save the relevant phenomena. The empiricist alternative in this case is implausible. The empiricist's description seems to require a dubious shift in the phenomenon being investigated. Psychometricians, those who are realists, take themselves to be investigating the development of psychological attributes; the models are purported to be models of attribute development. The empiricist, restricting herself to observable phenomena describes the situation quite differently, claiming that the shift to these new models is simply a move to models with greater empirical adequacy; the models are models of test data and the idiographic models are to be preferred because they fit the data better. The problem with this response is that the data sets used in traditional synchronic models are not the same as those used in the time-series models. To say that time-series models fit *the data* better than traditional models presupposes that there is a single set of data which the two approaches differentially represent, but this is not so. The idiographic turn is a shift not only in experimental design; it is also a shift in what is deemed empirically relevant data.

For the same reason that modeling covariation reflectively does not require a commitment to ontological realism, neither does the idiographic turn. That is, each factor structure at any point in a time-series may be considered a hypothesis subject to confirmation or refutation. Indeed the entire dynamical model of an individual's psychological change may be so considered. It would seem odd if one were propounding such hypotheses without any regard for their truth (or representational adequacy). The impetus for the turn is that time-series analysis offers better and more nuanced hypotheses, ones that are more likely to represent psychological processes. Nevertheless, the evidentiary requirements for ontological realism may remain unsatisfied despite best efforts, so ontological realism is not necessary to make sense of the idiographic turn.

5 Conclusion

Michell's admonishment that psychometricians conduct the relevant studies to ascertain the structure of attributes reveals Michell's commitment to MER while at the same time it suggests a methodological norm to the effect that ontological commitments need to be responsive to evidence. In none of the examples considered above has a case of ontological realism divorced from epistemic realism been encountered. In the case of model selection, we find evidence for methodological realism, but not necessarily a presupposition of ontological realism; in the case of using IRT models to predict item or test performance, no realist assumptions are required; in the case of using IRT models to estimate ability there is an implicit commitment to ontological realism, but that it is accompanied by a commitment to MER; in the assessment of the problem posed by non-locally homogeneous measurement model the commitment to ontological realism accompanies a commitment to MER; and in the case of the idiographic turn the development of statistical tools for the analysis of intraindividual variability evidences a commitment to MER, but not necessarily ontological realism. Psychometricians will have

various philosophical commitments, but there is nothing in the contemporary methods discussed here that requires pathological ontological commitment.

I have argued for methodological realism in psychometrics—that certain practices and methodological shifts in psychometrics seem to depend on realist commitments for their rationality. These commitments stand apart from the philosophical positions of particular psychometricians and even may be inconsistent with their professed commitments since background assumptions are not always clearly manifest, not even to those who harbor them. In some instances practice seems to indicate realist commitments where, upon closer inspection, there are none. In other instances close examination reveals a tacit commitment to ontological realism, epistemic realism, or both, though in none of the examples considered si there a commitment to ontological realism without a corresponding commitment to epistemic realism. Ontological realism unwarranted by evidence would provide a case for judging psychometrics as a "pathological science," and it was this pathology that concerned Gould and Michell. The charge that psychometrics is pathological finds no grounds in the examples provided. The assumption of epistemic realism is innocuous and represents a willingness to listen to evidence, i.e., a form of Longino's "responsiveness to criticism" criterion for scientific objectivity. Additionally, there are certain features of psychometric theorizing that cannot be explained on an empiricist account, and if attempting to make sense of science is one of the tasks of the philosophy of science, the empiricist is bound to fail in the case of psychometrics. Realism affords the epistemic, explanatory, and metaphysical richness needed to make sense of psychometric practice.

Acknowledgments I would like to thank Colin Allen, Denny Borsboom, Conor Dolan, Peter Jonason, Benjamin Lovett, and two anonymous reviewers for comments on earlier versions of this paper.

References

Baldwin, A. (1946). The study of individual personality by means of the intra individual correlation. *Journal of Personality, 14*, 151–168.

Bartholomew, D. (2004). *Measuring intelligence: Facts and fallacies*. New York: Cambridge University Press.

Boring, E. G. (1923). Intelligence as the tests test it. *New Republic, 35*, 35–37.

Borsboom, D. (2005). *Measuring the mind: Contemporary issues in psychometrics*. Cambridge: Cambridge University Press.

Borsboom, D., & Mellenbergh, G. (2004). Why psychometrics is not pathological: A comment on Michell. *Theory & Psychology, 14*, 105–120.

Borsboom, D., Mellenbergh, G., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review, 110*(2), 203–219.

Carnap, R. (1928). Der Logische Aufbau der Welt. Berlin: Weltkreis. Translated by R. A. George as *The Logical structure of the world and pseudoproblems in philosophy*. Berkeley: University of California Press, 1967.

Carroll, J. B. (1995). Reflections on Stephen Jay Gould's the mismeasure of man (1981). *Intelligence, 21*, 121–134.

Cervone, D. (2004). The architecture of personality. *Psychological Review, 111*, 183–204.

Coffa, J. A. (1991). *The semantic tradition from Kant to Carnap*. Cambridge: Cambridge University Press.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281–302.

Deary, I. J. (2000). *Looking down on human intelligence: From psychometrics to the brain.* Oxford: Oxford University Press.

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods, 5*, 155–174.

Ellis, J. L., & van den Wollenberg, A. L. (1993). Local homogeneity in latent trait models. A characterization of the homogeneous monotone IRT model. *Psychometrika, 58*(3), 417–429.

Fagan, M. B. (2007). *Objectivity in practice: Integrative social epistemology of scientific inquiry.* Dissertation, Department of History and Philosophy of Science, Bloomington: Indiana University.

Feigl, H. (1970). The "orthodox" view of theories: Remarks in defense as well as critique. In M. Radner & S. Winokur (Eds.), *Analyses of theories and methods of physics and psychology. Minnesota studies in the philosophy of science Vol IV* (pp. 3–16). Minneapolis: University of Minnesota Press.

Friedman, M. (1999). *Reconsidering logical positivism.* Cambridge: Cambridge University Press.

Gould, S. J. (1996). *The mismeasure of man.* New York: W.W.Norton & Company. Original edition, 1981.

Hamaker, E. L. (2004). *Time series analysis and the individual as the unit of psychological research.* Dissertation, Department of Psychological Methods Amsterdam: University of Amsterdam.

Hambleton, R. K., Swaminathan, H., & Rogers, J. (1991). Fundamentals of item response theory. In R. Jaeger (Ed.), *Measurement methods for the social sciences.* Newberry Park: Sage.

Hilbert, D. (1899). *Grundlagen der Geometrie.* Leizpig: Teubner. Translated by L. Unger from the 10th edition as *Foundations of Geometry.* La Salle: Open Court, 1971.

Hood, S. B. (2008). Comments on borsboom's typology of measurement theoretic variables and Michell's assessment of psychometrics as pathological science. *Measurement Interdisciplinary Research and Perspectives, 6*, 93–96.

Jensen, A., & Weng, L. (1994). What is a good *g*? *Intelligence, 18*(1), 231–258.

Kelley, T. L. (1927). *Interpretation of educational measurements.* New York: Macmillan.

Kitcher, P. (2001). *Science, truth, and democracy.* Oxford: Oxford University Press.

Leplin, J. (1986). Methodological realism and scientific rationality. *Philosophy of Science, 53*, 31–51.

Leplin, J. (1997). *A novel defense of scientific realism.* New York: Oxford University Press.

Lloyd, E. A. (2005). *The case of the female orgasm: Bias in the science of evolution.* Cambridge: Harvard University Press.

Longino, H. E. (1990). *Science as social knowledge.* Princeton: Princeton University Press.

Longino, H. E. (2002). *The fate of knowledge.* Princeton: Princeton University Press.

McMullin, E. (1983). "Values in science", In P. Asquith, & T. Nickles (Eds.), *PSA 1982*(2), 3–28. East Lansing: Philosophy of Science Association.

McMullin, E. (1987). Explanatory success and the truth of theory. In N. Rescher (Ed.), *Scientific inquiry in philosophical perspective.* Lanham: University Press of America.

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: A critical history of a methodological concept.* Cambridge: Cambridge University Press.

Michell, J. (2000). Normal science, pathological science and psychometrics. *Theory & Psychology, 10*, 639–667.

Michell, J. (2004). Item response models, pathological science and the shape of error. *Theory & Psychology, 14*, 121–129.

Michell, J. (2008). Is psychometrics pathological science? *Measurement Interdisciplinary Research and Perspectives, 6*, 7–24.

Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika, 50*, 181–202.

Molenaar, P. C. M. (1999). Longitudinal analysis. In H. J. Ader & G. Mellenbergh (Eds.), *Research methodology in the social, behavioural, and life sciences* (pp. 143–167). London: Sage.

Molenaar, P. C. M., Huizenga, H. M., & Nesselroade, J. R. (2003). The relationship between the structure of inter-individual and intra-individual variability: A theoretical and empirical vindication of developmental systems theory. In U. M. Staudinger & U. Lindenberger (Eds.), *Understanding human development.* Dordrecht: Kluwer.

Nyborg, H. (2003). The sociology of psychometric and bio-behavioral sciences: A case study of destructive social reductionism and collective fraud in 20th century academia. In N. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur Jensen* (pp. 441–502). Amsterdam: Pergamon.

Psillos, S. (1999). *Scientific realism: How science tracks truth*. New York: Routledge.
Schlick, M. (1918). *Allgemeine Erkenntnislehre*. Berlin: Naturwissenschaftliche Monographien und Lehrbücher. Translated by A. Blumberg as *General theory of knowledge*. La Salle: Open Court, 1985.
Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology, 5*, 201–293.
Spearman, C. (1927). *The abilities of man: Their nature and measurement*. New York: Macmillan.