ARTICLE IN PRESS

Measurement xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

On the philosophical foundations of psychological measurement

Andrew Maul [a,*], David Torres Irribarra [b], Mark Wilson [c]

[a] Gevirtz Graduate School of Education, University of California, Santa Barbara, United States
[b] MIDE, Pontificia Universidad Católica de Chile, Chile
[c] Graduate School of Education, University of California, Berkeley, United States

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Educational and psychological measurement
Philosophy of measurement
Philosophy of science
Empiricism
Pragmatism
Scientific realism

ABSTRACT

Measurement has played a central role in the development of the physical sciences and engineering, and is considered by many to be a privileged method for acquiring information about the world. It is thus unsurprising that the psychological sciences have also attempted to develop methods for measurement. However, it is not clear how the ways in which psychological scientists understand measurement accord with how the concept is understood in other scientific disciplines, or by the professional and general publics. In part this may be due to the ways in which several distinct strands of thinking about scientific inquiry (and measurement in particular) have influenced the work of psychological scientists over roughly the past hundred years. Given that such influences are often not studied or even acknowledged, many psychological scientists may be unaware of the resulting tensions in their conceptual vocabulary, and of the gaps between the nature of their claims on psychological measurement and the substantiation for those claims. The aim of this paper is to overview the major philosophical influences on thinking about psychological measurement, and to note the pitfalls of some of the extreme positions that have emerged. We hope that such an overview may help facilitate greater clarity concerning the semantics of measurement claims made by psychological scientists.

1. Introduction

Measurement has long been an important and prominent concept in the physical sciences, engineering, and natural philosophy, and is often considered a privileged method for acquiring information about the world (e.g., [38]). Given this, it is unsurprising that the psychological sciences[1] have, since their inception, developed a variety of techniques that purport to be instances of measurement as well [23,45]. However, it is not clear how the ways in which psychological scientists understand the concept of measurement accord with how measurement is understood in other scientific disciplines, or by the professional and general publics.

An obvious difference between the psychological and physical sciences concerns the nature of the attributes[2] that commonly come under investigation in each of these fields. In the psychological sciences it is common to hear claims on the measurement of sociological attributes such as 'cultural capital' and 'socio-economic status', psychological attributes such as 'anxiety' and 'self-esteem', and more

* Corresponding author.
E-mail address: amaul@education.ucsb.edu (A. Maul).

[1] The term "psychological sciences" refers here to all scientific disciplines and activities concerned with gaining knowledge of the human mind and behaviour, including not only psychology, but also sociology, anthropology, and disciplines of research concerned with particular human activities such as education, political science, and industrial organizations. Thus, the term is interpreted analogously with the term "physical sciences," which refers not only to physics but also other disciplines concerned with physical material, such as chemistry, biology, geology, and astronomy.

[2] We use the term "attribute" to refer both to what are sometimes called "properties" (e.g., mass) and "relations" (e.g., weight, which is a relation between mass and local gravity).

http://dx.doi.org/10.1016/j.measurement.2015.11.001
0263-2241/

classically academic attributes such as 'mathematical proficiency' and 'college readiness'. *Prima facie*, such attributes appear to be significantly different from physical attributes such as spatial distance and temperature: in particular, psychological attributes would seem to be far less likely to show invariant relations with other attributes or to operate causally in networks of laws, due to the ways in which these kinds of attributes are dynamically indexed to particular cultural, social, and historical conditions. Additionally and relatedly, there is far less agreement amongst psychological scientists concerning the meaning of psychological concepts than there is amongst physical scientists concerning (most) physical concepts; even high-profile attributes such as 'intelligence' and 'depression' remain controversial. The types of knowledge, skills, abilities, and other personal attributes (collectively, KSAs) targeted by educational programs may seem even less tractable, given that such attributes are, to an important extent, defined by socially, culturally, and historically situated perspectives and concerns, as well as current theories of cognition. The socially dependent (one might say "constructed") nature of such attributes is at the centre of objections that some traditional understandings of measurement present to the use of the concept in psychological sciences. How can an attribute that is constructed by humans be a quantity, or a *real* property at all? Even if one accepts that such attributes can be modelled as quantities, they are surely resistant to standard techniques of (physical) empirical manipulation such as concatenation, which arguably eliminates them as candidates for 'fundamental' measurement ([16]; also see [54, p. 186]); if instead we claim to be able to evaluate their structure indirectly (e.g., via additive conjoint measurement; [33]), how can we deal with the measurement error present in nearly all psychological applications [19]; also see [8], ch. 4?

Finkelstein (e.g., [24,25]) drew a relevant distinction between the measurement of "hard" and "soft" systems, describing the latter in terms of domains that involve "human action, perception, feeling, decisions and the like" [25, p. 269], and noting that invariant relations could likely not be established amongst "soft" systems due to the absence of "adequately complete" and validated theories. A variety of sub-fields in the psychological sciences (including psychometrics, econometrics, mathematical psychology, and psychophysics) have developed techniques that purport to permit the measurement of attributes in "soft" systems, but the claims made in these subfields remain controversial; notably, in recent years, a number of scholars (e.g., [8–10,12,13,11,19,28,34–36,41,43–47,49,50,63,65,67,73,77,78]) have subjected the conceptual and philosophical foundations of psychological measurement to vigorous investigation and critique. The conclusions of these inquiries have often turned up unfavourably regarding both the actuality and even the possibility of psychological measurement (for the former, see in particular [44,45,46,47,49,50]; for the latter, see [77,78]), evidencing that the way in which measurement is understood by psychological scientists may be entirely dissimilar to the way in which it is understood by physical scientists and philosophers of science. Accordingly, at least by some traditional criteria, there may not have yet been any instances of successful measurement of psychological attributes.

In our experience, most members of the mainstream educational and psychological measurement and assessment community are simply unaware of this body of work, as well as the literature on metrology and the history and philosophy of measurement more broadly; further, those who are tend to react dismissively. To the extent to which such dismissals are made explicit, most can be characterised in one of two distinct ways. The first type of response involves acceptance of claims made by scholars such as Michell [92] that "within scientific contexts the term measurement has only *one* meaning and that is as the *assessment of quantity*" (p. 127, emphasis in original) – or at least that there are certain essential features of measurement that may be shared amongst somewhat different instantiations – and that the activities and conceptual vocabulary of psychological scientists are inconsistent with this definition or these essential features. In this case, the concept of 'measurement', when used by psychological scientists, would be seen as a metaphor at best (see, e.g., [52]) and a conceptual error at worst. The second type of response involves denial of the premise that the concept of measurement has or needs to have a consistent definition (or even common essential elements) across scientific disciplines; it is thus concluded that psychological scientists and physical scientists may unproblematically maintain entirely dissimilar understandings of measurement. In this case, the word "measurement", when used by psychological scientists, would be merely a homonym for the word used in other disciplines. In principle, this type of response would need to be accompanied by an alternative account of how measurement concepts are to be understood, especially to the extent to which psychological scientists continue to engage in practices that make use of the logic and vocabulary of classical measurement (as will be exemplified further in later sections); in our experience, however, such an alternative account is generally not given, leaving the concept of psychological measurement and its relation to other forms of measurement nebulous at best. In both types of response there is an implied rejection of the idea that success in the psychological sciences depends on – or, possibly, could even benefit from – conceptualising measurement in a manner consistent with its historical usage in other scientific and philosophic disciplines; this may be associated with a broader rejection of the idea that it is desirable (or, perhaps, possible) for different scientific fields to have common understandings of the practice of measurement.

This largely apathetic or dismissive attitude regarding the meaning of measurement may relate to entrenched traditions in the training of psychological scientists (c.f. [84,79]), which typically contains very little or no direct instruction on the historical and conceptual foundations of measurement. A telling illustration of this point was provided by Borsboom [10], in his review of the most recent edition of *Educational Measurement*, a heavily-cited and highly-regarded volume which proclaims on its own back cover to be "the bible in its field," in which he noted that "although the word *measurement* figures as prominently in the book as the title suggests, there is no discussion of what it might mean; no discussion of the extant philosophy of science literature on the topic; no discussion of formal

Please cite this article in press as: A. Maul et al., On the philosophical foundations of psychological measurement, Measurement (2015), http://dx.doi.org/10.1016/j.measurement.2015.11.001

measurement theory; no discussion of how the activity of measurement relates to the activity of testing; and no discussion of the relevance of all these issues to the question of validity and the process of validation... [most of] the authors of the book chapters appear to consistently conflate the terms *testing* and *measurement*, as if any test score is automatically a measure, any testing procedure is automatically a measurement procedure, and test theory is more or less the same thing as measurement theory ... [which] is clearly mistaken" (p. 706, emphasis in original).

Concerns such as these lead some proponents of a traditional understanding of measurement to conclude that the psychological sciences may not be well suited for rigorous measurement; the situation is surely not aided by the fact that psychological scientists generally seem uninterested in addressing such issues, tacitly adopting instead the perspective that virtually anything can be measured (sometimes adding the caveat that measurement is achieved if particular statistical criteria are met; see [8] for a discussion of some such positions).

At least in part, this situation may be due to the ways in which several distinct strands of thinking about measurement have influenced the conceptual vocabulary of psychological scientists over roughly the past hundred years.[3] In this paper, we review and discuss what we consider to be the most influential of these strands, noting in particular the pitfalls that may be associated with over- or mis-interpretation of each of them. While acknowledging that any attempt at categorisation risks oversimplification, we group these strands as follows: first, the *neo-positivist* (and, in particular, *operationalist*) accounts of measurement emerging from the early-to-mid twentieth century, second, the influences of *pragmatist* philosophy on thinking about educational and psychological measurement, especially in the later twentieth century and early twenty-first century, and third, *realist* thinking about measurement, which has been most explicit in relatively recent writing by scholars such as Joel Michell, but is arguably implicit in a great deal of other work as well. In the final section, we briefly discuss possibilities for a reconciliation of some of the moderate elements of each strand of thought.

2. Neo-positivism, operationalism, and psychological measurement

Broadly, most of Western science is in some sense *empiricist*, in the sense that direct observation is taken as the basis for knowledge (although what counts as direct observation is itself an unsettled issue; see [6]). Standard accounts of the history of Western science associate the origins of empiricism with Greek and Persian antiquity (for example, in the writings of Aristotle, Al Farabi, and Avicenna), and its maturity with the Italian Renaissance (typified by the work of Galileo) and the British Enlightenment (especially, the writings of Bacon, Hobbes, Locke, and Hume).

In the twentieth century, the movement known as *logical positivism* synthesised many ideas from classical empiricism along with then-current advances in the philosophy of language and mathematics. Logical positivism was associated with the position that statements regarding unobservable (theoretical) entities or forces should only be regarded as meaningful if such statements can be linked to observations in a clear and consistent manner. In the fledging psychological sciences, *behaviourism* (e.g., [69]) captured many of the same intuitions as those behind positivism, such as an emphasis on observables as the basis for science and an imperative to avoid metaphysical theories and concepts. In particular, the concept of the human mind was regarded as too metaphysical and unobservable to be a proper object of scientific inquiry.

The positivist project is now widely regarded to have failed, and stricter forms of behaviourism (especially as a logical doctrine, as distinct from a methodological doctrine) have been abandoned [3]; modern philosophies of science typically include a much greater degree of acceptance of the inclusion of unobservable phenomena – such as the human mind – in scientific theories. However, positivism and behaviourism have left a significant legacy on methodological thinking in the psychological sciences, including thinking about measurement [26,3]. This is perhaps most visible in the influence of *operationalism*, which is consistent with many positivist principles that emerged in the early-to-mid 20th century. Operationalism (or operationism) was originally articulated by Bridgman [15] as the thesis that "we mean by any concept nothing more than a set of operations; *the concept is synonymous with a corresponding set of operations* (p. 5, emphasis in original). Operationalism was originally proposed as a semantic doctrine about the meaning of theoretical terms rather than a theory of measurement *per se*: operationalism holds that the meaning of theoretical terms is exhausted by the particular operations undertaken to observe them, which means that the results of a particular set of operations (or measurement procedure) are interpreted as measurements by fiat.

Operationalism had a strong influence on psychology through the influence of early behaviourists such as Skinner [68], Boring (e.g., [7]), and his student Stevens (e.g., [70]; see [86,26,41]). One obvious reason for the attractiveness of operationalism to early psychologists is the difficulty of precisely defining psychological attributes (which, as discussed earlier, are not 'observable' by traditional criteria, and seemed dangerously metaphysical in contrast to the positivist and behaviourist zeitgeist of the time); stating, for example, that "intelligence is what the tests test" [7, p. 35] neatly sidesteps the issue, and also gives at least the appearance of rigor and objectivity by reducing abstract ideas to observables. Notable in the concept of operationalism is that "it is meaningless to ask whether something is 'really' being measured... there is neither a need nor a place for postulating attributes which are prior to the measurement operation" [8, p. 93]. The most widely-used definition of measurement in the psychological sciences (see [44], for a defence of this claim), namely Stevens' view that "measurement, in the broadest sense, is defined as the assignment of numerals to objects

[3] This paper does not aim to provide a comprehensive survey of all philosophical stances on measurement (though see [74]); rather, we focus here on the lines of thinking that have been most strongly influential in the psychological sciences.

Please cite this article in press as: A. Maul et al., On the philosophical foundations of psychological measurement, Measurement (2015), http://dx.doi.org/10.1016/j.measurement.2015.11.001

or events according to rules" [70, p. 667], is consistent with operationalism, insofar as, taken at face value, this statement implies that the only necessary condition for measurement is the presence of a rule (operation) for numerical assignment. Stevens was quite clear that this could be *any* rule other than "random assignment, for randomness amounts in effect to a nonrule" [71, p. 47].

This understanding of operationalism then makes it exceedingly easy to measure nearly any psychological attribute. For example, in an educational context, any form of knowledge, skill, or ability (e.g., 'mathematical proficiency') can be measured simply by assembling a set of questions that are judged (by whatever criteria) to be relevant to the specified KSAs; the attribute of mathematical proficiency could then be defined in terms of the sum of the number (or percentage) of correct answers a student gives to the set of questions. On this understanding, the question of whether measurement is 'actually' taking place may seem odd or ill-formed, helping explain some of the previously-described apathy and resistance of psychological scientists to conceptual challenges related to measurement.

But despite the attractive simplicity of this option, difficulties immediately present themselves. Most obviously, such an interpretation is out of step with common vocabulary and the logic of education: most educators would immediately recognise that students' knowledge, skills, and abilities are not *equivalent* to their score on a particular test, and that such an identification narrows the definition of the attribute of interest to what usually would be considered just an *indicator* of the relevant KSAs. Further, it may invite the misconception that properties of the numerical assignment or total score are identical to properties of the measured attribute, "so that, for instance, attributes are presumed to induce a linear ordering of people because sumscores do" [79, p. 429].

Operationalism – at least in its original formulation[4] – has been almost uniformly rejected as irreconcilable with general scientific practice and vocabulary [27,3]. One reason for this is that operationalism has the consequence that each unique set of operations must be associated with a distinct theoretical term; thus, for example, the outcome of the application of an alcohol thermometer and the outcome of the application of a mercury thermometer cannot refer to the same theoretical property, nor could two distinct tests of mathematical proficiency be claimed to measure the same attribute [8,27]. Additionally, the related concepts of measurement error and measurement uncertainty – the former a central concept in psychological measurement, and both central concepts in present-day metrology [30] – are ill-fitting with operationalism: if the results of applying a procedure are by definition a measurement of the theoretical term, what is there to be in error about?

Following the collapse of logical positivism and an associated general retreat from the more extreme forms of empiricism, many scholars became increasingly willing to accept that the interpretation of concepts like temperature and knowledge outrun their associated measurement procedures – that is, theoretical concepts are seldom exhausted by their operational definitions – and, in fact, it is very difficult to make sense of both scientific and lay discourse about such concepts without this belief. More broadly, there is a distinction between the subject matter of a scientific discipline (i.e., its ontology) and its methods of acquiring knowledge about its subject matter (i.e., its epistemology); as was argued by Chomsky [18], psychology is not the science of behaviour any more than physics is the science of metre readings (see [51]).

The influence of behaviourism and other strong forms of empiricism on psychological scientists' thinking about measurement was not limited to operationalism. Mid-twentieth-century authors writing on the topic of validity in psychological testing such as Cureton [22], Cronbach and Meehl [21], and Campbell and Fiske [82] were heavily influenced by logical positivism and logical empiricism as well [87]. One way in which this is visible is the focus of these authors on correlations between test scores and other outcomes as the primary determinant of validity, and even as a tool for fixing the identity of a 'hypothetical construct' [21]; c.f. [12]. In a related but distinct line of scholarship, scholars such as Cureton [22], and, more recently, Cronbach et al. [20], have invoked behavioural domain theory to help fix the identity of measured attributes, either by describing attributes in terms of domains of behaviour from which observed behaviours are taken to have been sampled (e.g., [22]), or as 'dispositional' attributes defined in terms of their possible sets of behavioural consequences (e.g., [20]). Such attributes are often interpreted instrumentally, rather than realistically; for example, Lord et al. [37] note that "nowhere is there any necessary implication that traits *exist*... it is sufficient that a person behave as if he were in possession of a certain amount of ... relevant traits and that he behave *as if* these amounts substantially determined his behaviour" (p. 358, emphasis added).

A final example of the influence of stronger forms of empiricism on thinking about measurement in the psychological sciences can be found in the literature on representational measurement theory (RMT), which is characterised by the stance that measurement is the construction of a representation of an empirical relational system via a numerical relational system (e.g., [33,54]). On this view, the starting point for measurement is the determination of empirical relations amongst objects (e.g., X is greater than Y and less than Z). Consistently with positivist principles, this requires that empirical relations be directly observable, or "identifiable" [72, p. 7], though it is not always obvious what this means (c.f. [8,43]).

Once empirical relations are determined, numbers are assigned to empirical entities in such a way as to preserve the qualities of their empirical relations. Relational systems can possess different sorts of structures, and the particular sort of mapping of empirical onto numerical relations determines the scale properties. Many psychological scientists are familiar with this logic through exposure to Stevens' [70] 'levels of measurement' (nominal, ordinal, interval, ratio), which are scale types defined by their

[4] Bridgman [81] later revised and softened his initial position (see also [17]); unfortunately, there does not appear to be any evidence that this softening impacted the thinking of psychological scientists. Also, it has been argued that Stevens himself probably did not endorse the early, more extreme version of Bridgman's operationalism (Feest, 2005, cited in [74]) and may have even had realist inclinations.

Please cite this article in press as: A. Maul et al., On the philosophical foundations of psychological measurement, Measurement (2015), http://dx.doi.org/10.1016/j.measurement.2015.11.001

'admissible transformations' (i.e., the ways in which the numerical relations can be manipulated while maintaining their homomorphisms to specified empirical relations). However, apart from RMT's influence on Stevens, its impact on the work of psychological scientists has been very limited [19]. In part this seems to be because RMT, in its unadorned form, cannot account for the role of measurement error except via the introduction of some version of scientific realism [8]. Thus, it is very difficult to make sense of how RMT accords with both scientific and lay discourse about measurement and knowledge acquisition more generally [43]. Although there have been attempts to reconcile RMT with probabilistic approaches (e.g., [14,88]; cf. [34]), these attempts have not, so far, led to a change in RMT.

In sum, measurement, on the broadly empiricist view, is about the connection between specific (*actual*) observations and the (usually numerical) outcomes of a specified procedure. In operationalism, the relationship is one of fiat; in some other interpretations, the relationship may be one of inductive summaries of sampled behaviours or instrumental devices for prediction. At least on the strongest interpretations of these empiricist views, there is no need to represent anything else – such as unobserved *causes* of behaviour – in a depiction of what measurement is all about.

3. Pragmatism and psychological measurement

The origins of the philosophical movement known as *pragmatism* are generally associated with Pierce (e.g., [56]) and James [29]. Although there are distinctions between their formulations, and perhaps even greater distinctions between the ways in which pragmatist philosophies have been formulated by more recent scholars such as Rorty et al. (e.g., [62]), Putnam (e.g., 1999), and Toulmin [76], pragmatist orientations generally share in common an orientation towards action (i.e., practice) and utility. In the words of James [29]:

> "The pragmatic method ... is to try to interpret each notion by tracing its respective practical consequences. What difference would it practically make to anyone if this notion rather than that notion were true? If no practical difference whatever can be traced, then the alternatives mean practically the same thing, and all dispute is idle" (p. 14).

From the pragmatist's perspective, the purpose of beliefs and theories are to facilitate successful engagement with the world, rather than (necessarily) to describe the world as it truly is [4]; thus, beliefs and theories are most appropriately judged by their usefulness in facilitating such successful engagement [61,5]. Thus, a pragmatic orientation towards measurement would likely de-emphasise concerns such as the ontological status of the measured attribute or the manner in which numerical assignments are formally constructed out of empirical relations, and ask instead whether and how the results of the measurement procedure can be usefully put to work to achieve a particular set of goals. The concept of 'usefulness' is, of course, relative to the goals, motivations, and values of individuals, groups, and society as a whole, and thus as such motivations and values change, so might the perceived usefulness of the measure.

All this stated, though references to 'pragmatism' abound in the literature on educational and psychological measurement, connections to the larger pragmatic philosophical tradition are rarely made explicit. Put simply, when psychological scientists speak of a *pragmatic point of view* or a *pragmatic approach*, they often seem to be invoking a common-language interpretation of the term, meaning something like *concerned with or relating to matters of fact or practical affairs; practical rather than idealistic or theoretical.*

Accordingly, although there are few if any explicitly pragmatic theories of measurement[5] within the psychological sciences, and especially within areas of applied scholarship such as education, a focus on the practical consequences of theories and methodologies – a prominent aspect of the pragmatic tradition – has gained considerable popularity. For example, the concept of *validity* is defined by Messick [42] as a "judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment" (p. 13, emphasis original). The focus on *actions* taken on the basis of tests, in particular, is a common theme throughout much of the literature on validity (e.g., [31,2,66,91]); it could also be noted that this definition of validity contains no reference to *truth per se* (c.f., [11]). The "adequacy and appropriateness" of test scores are then evaluated using Toulminian models of evidentiary argumentation [31,52]. As Kane [31] (p. 60) states, "the argument-based model provides a relatively pragmatic approach to validation. The goal is to develop a measurement procedure that supports the proposed interpretations and uses and an... argument that is plausible, given the measurement procedure."

The foregrounding of usefulness rather than truth as the primary criteria for success in psychological measurement can be traced at least as far back as Lord et al. [37], in which (in the context of discussing the criteria for what counts as an 'interval' scale), they assert that "*from a pragmatic point of view*, the only meaningful evaluation of this procedure is one based on an evaluation of the usefulness of the resulting scale" (p. 22, emphasis in original). Prima facie, such a focus seems sensible – perhaps even inevitable – insofar as psychological and educational tests and assessments are generally developed with a certain purpose or set of purposes in mind and are deeply embedded in complex and dynamic social structures. But although a pragmatic focus on formulating and defending arguments for proposed uses of tests may initially seem straightforward,

[5] Exceptions include the work of Adams [1], who offers a perspective on measurement in which goals are foregrounded as a central element, and Torres Irribarra [75], who presents a recent effort to formulate a perspective on measurement thoroughly rooted in the tenets of philosophical pragmatism. Additionally, authors may sometimes invoke pragmatist principles to solve particular problems without explicitly committing to a pragmatic theory of measurement; a notable example is Sherry's [89] treatment of the justification of claims regarding quantities and quantification.

Please cite this article in press as: A. Maul et al., On the philosophical foundations of psychological measurement, Measurement (2015), http://dx.doi.org/10.1016/j.measurement.2015.11.001

the actual work of unpacking what forms of empirical evidence and theoretical rationales are necessary for the defence is seldom so. In part this is because no account is typically given of what knowledge is necessary – and thus, what forms of evidentiary substantiation are needed – for the appropriate use of test scores for particular purposes.

Additionally, and more broadly, whereas most philosophical traditions understand the concept of measurement as an enterprise aimed at the acquisition of (usually some specific form of) knowledge (see, e.g., [73]), on the pragmatic philosophical tradition the acquisition of knowledge is not seen as fundamentally distinct from any other use. As far as measurement goes, at least two claims about knowledge could be formulated from a purely pragmatic perspective. The first is that, since testing procedures are to be judged on the criteria of utility rather than truth, the question of whether or not any actual knowledge-acquisition is taking place is no longer a critical issue. The second is that, as with any other concept, *all that is meant* by 'measurement' in any given setting is contained in the uses and practical consequences of the assessment procedure and its outcomes. The first claim would allow for a distinction to be made between *assessment* (or testing in a more narrow sense) as a socially situated activity with a particular set of aims, and *measurement*, understood as a knowledge-acquisition enterprise, but would regard the latter as unimportant except insofar as it informs the former; thus, for example, measurement may still be regarded as an enterprise that depends on a connection to independently-existing attributes, but that is of (at most) secondary importance. The second claim is most consistent with the Jamesian definition of truth as utility; according to this view, 'measurement' could be considered to be synonymous with 'testing.'

However, despite the popularity of pragmatism (in the narrow sense) with the educational and psychological measurement communities, much of the time it does not seem to be the case that either of these claims (or any other coherent claim) is being formally made:[6] rather, practicality and expedience often seem to be invoked as an excuse to not examine or recognise deeper conceptual issues. Much of the time it is not clear what conceptions of either measurement or usefulness (i.e. a critical component of a pragmatist approach) are being invoked in discussions of testing, and what the relationship between measurement and testing is taken to be; as is well illustrated by Borsboom's [10] previously-cited review of *Educational Measurement*, much of the time it seems that testing is simply equated with measurement, which we agree "is clearly mistaken" (p. 706).

To see more clearly why this is mistaken, one must only note that tests can have many uses that do not involve anything that would be conventionally recognised as measurement, and in fact can be successful by many important criteria without measurement ever taking place. For example, some tests may be successfully used as deterrents for certain behaviours (e.g., a company-mandated drug test or an airport security screening) or prods to certain forms of action (e.g., if an upcoming fitness test inspires a person to exercise), even if the tests in question fail to measure anything at all.

More commonly, and perhaps insidiously, tests may simultaneously have both measurement and non-measurement goals. The failure to distinguish measurement from testing more broadly is especially problematic when the measurement goals of testing are not recognised as being in need of evaluation and justification in their own right. The lack of a distinction between these concepts makes it difficult to recognise when genuine claims about the measurement of psychological attributes are implicit in larger claims about the use of tests. For example, it may be claimed that it is appropriate to base decisions about the retention or firing of teachers on the results of educational tests given to their students. It is very difficult to make sense of this claim without additional claims such as (a) educational tests provide information about important attributes of students, as well as (b) teachers have causal agency on these student attributes, and (c) there are adequate ways to account for other causal influences on student attributes. Further, given that the mode of expression of the information obtained is numerical (and, more specifically, is almost invariably presented in terms of continua), one is strongly tempted to rephrase these goals in terms of measurement: educational tests measure important attributes of students, and thus, given (b), these tests measure teacher effectiveness. Articulation and explicit evaluation of such measurement claims is completely missing from many discussions of the validity of the use of educational tests for the purposes of teacher evaluation.[7]

Thus, a great deal of thinking about psychological measurement is characterised by the pragmatism-influenced view that tests should be evaluated with reference to their intended purposes, which we take to be reasonable. However, appealing to a narrow understanding of pragmatism to contend that testing is synonymous with measurement,

[6] We are unaware of anyone in the educational and psychological measurement and assessment community who has committed themselves to the second position, i.e., that the entirety of the meaning of measurement is contained in the practical consequences of the testing procedure.

[7] On occasion we have heard the suggestion that this whole conversation could be avoided if the educational and psychological science communities would simply abandon usage of the term 'measurement', and instead exclusively use terms such as 'testing' or 'assessment.' We are not optimistic about the promise of such a lexical switch, for at least two reasons. First, the use of 'measurement' is not accidental, nor trivially identical to 'testing' or 'assessment' in the minds of either scientists or laypersons: as previously noted, the concept of measurement has a long and prestigious history of usage in the physical sciences, and as such commands a great deal of social capital; it could even be argued that many scientists and laypersons see measurement as a necessary component of scientific inquiry [45]. The most well-known example of measurement, namely the measurement of spatial distance, is frequently used analogically by educators and psychological scientists, further encouraging the perception that measurement in the psychological sciences is directly analogous to measurement in the physical sciences. Second, a great deal of the conceptual vocabulary surrounding putative measurement in the psychological sciences is parasitic on the manner in which measurement is understood in the physical sciences [12,13,41,44,45]; that is, it is not simply the *word* 'measurement' that has been imported from the physical sciences, but also foundational concepts such as quantity, causality, attributes (or properties), units, ratios, systems of lawful relationships, and so forth. We are not aware of any attempts to articulate the semantics of testing in a way that does not depend on measurement concepts.

Please cite this article in press as: A. Maul et al., On the philosophical foundations of psychological measurement, Measurement (2015), http://dx.doi.org/10.1016/j.measurement.2015.11.001

or to avoid examining the distinction, seems profoundly misleading, and surely contributes to misunderstanding amongst the public regarding their origins and interpretations.

4. Realism and psychological measurement

As with empiricism and pragmatism, there is a wide range of positions that fall under the broad headings of *realism* and *scientific realism*. Broadly, they share in common some variant of the belief that at least one of the aims of inquiry is to acquire knowledge about reality, which is taken to be in at least some sense independent of the inquirer. On the realist's view, the (perhaps approximate) *truth* of scientific theories is generally taken to be the best explanation for their success. Different varieties of realism may or may not endorse the commitments that (a) there is a (single) natural world, which exists regardless of what any conscious being thinks or perceives (the 'metaphysical commitment'); (b) claims about the world are to be taken at face value, as possessing truth-values (the 'semantic commitment'), and (c) so interpreted, true claims constitute knowledge of the world (the 'epistemological commitment'). Also, realism may be applied to entities (e.g., electrons), attributes (e.g., electrical charge), relations (e.g., the ratio between the electrical charge of a given object and the coulomb), and theories (e.g., that there is a causal relationship between electrical charge and the readings of an electrometer). In general, realists about measurement hold that measurement is aimed at the acquisition of knowledge about attributes, where the attributes in question are held to exist independently of the specific measurement procedure and the language, thoughts, and conventions of the persons performing the measurement.

Realism may be regarded as something of a default position amongst practicing scientists and laypersons alike [83], and this is no less true in the psychological sciences than the physical sciences ([84], preface). Although there have been a few explicit treatments of realism in psychological measurement in relatively recent years (e.g., [43,90]), a great deal more of the thinking in this area could be argued to *implicitly* invoke some form of realist commitment or assumption. For example, the early (and still very common) idea that a test is valid if it measures what it intends to measure [85] assigns an independent identity to the measured attribute. Also, claims regarding concepts such as estimation and measurement error are (prima facie) difficult to make sense of without the belief that there is a true value of an attribute to be in error about (fallibilism). Additionally, it is difficult if not impossible to coherently interpret the semantics of latent variable models, such as are commonly used in psychological and educational research, without realist semantics [12,9].

However, despite the ubiquity of realist ideas about psychological measurement, many psychological scientists nonetheless resist explicitly realist formulations of measurement claims, as evidenced by the relative dearth of formal accounts of realism in the psychological and educational literature. One possible reason for this is that realism is sometimes misperceived as necessarily entailing a commitment to some variant of the idea of 'absolute truth,' or the position that there exists one true and complete description of the way the world really is, and that inquiry itself will be complete once such a description is in hand. This position is sometimes associated with an unflattering portrait of logical positivism or other forms of empiricism, and contrasted with other positions such as social constructionism or postmodernism (see [84], preface). In the context of psychological measurement, the implication that measured attributes of objects must 'objectively' exist in order to be measurable is often interpreted as implying that they must possess physical (perhaps specifically neurophysiological) identity, and in some contexts a genetically-determined biological basis for variation in the attribute; such claims may evoke a negative reaction from many scholars familiar with, for example, the controversial history of intelligence testing and its association with race (see [55], for a recent review).

Of course, realism in general need not be associated with this strong form of 'naïve' realism, nor do most philosophers formulate realist claims in this way. Putnam (e.g., [57–59]), for example, argues that there are simply too many ways in which beliefs and symbols can be mapped onto the world for it to be plausible that there could be a single best description of the way the world is. A famous, though not uncontroversial, example of this is the observation that "theories of space–time can be formulated in one of two mathematically equivalent ways: as an ontology of points, with spatiotemporal regions being defined as sets of points; or as an ontology of regions, with points being defined as convergent sets of regions. Such theories are descriptively equivalent since mathematically equivalent and yet are logically incompatible from the [naïve] realist's point of view" [32]. Somewhat less formally, and in terms of human psychology, it is often the case that there are many possible ways to describe psychological phenomena that are equally consistent with all available empirical data, but are semantically inconsistent (e.g., [53], cited in [8]). As a separate but related point, it would seem to be a fairly straightforward observation that the meaning of terms about human beings, both in informal and formal discourse, is indexed to particular sociohistorical conditions; for example, the meaning of a term such as 'nursing competence' is likely to change over time (as new medical technologies are developed, roles of hospital staff change, etc.), and from one geographical region to another, and even from one hospital ward to another (see [40]). Scholars such as Messick [42] and Mislevy [52] have formulated versions of 'constructive-realism' that allow for the idea that attributes measured by educational and psychological tests are, to an important extent, defined by socially-, culturally-, and historically-situated perspectives and concerns, as well as current theories of cognition, all of which may vary over time, and between different stakeholders at any given time. An example of a philosophic framework that is broadly consistent with such a view is found in Putnam's recent (e.g., [60]) writings on *pragmatic realism*, which acknowledge that conceptual pluralism is not at odds with realism, but rather, "to use a Wittgensteinian idiom, *seeing* is always *seeing as*, and it is

Please cite this article in press as: A. Maul et al., On the philosophical foundations of psychological measurement, Measurement (2015), http://dx.doi.org/10.1016/j.measurement.2015.11.001

the interface between the world and the rich fabric of our concepts that jointly determines what we see" [60, p. 20]. On such an account, the existence of natural ('objective') reality is not denied, but neither is it seen as directly presented to our senses; instead, our conceptual schemes, models, and linguistic frameworks actively shape our experiences and frame our knowledge of the world, causing us to organise and prioritize experience in a particular way, leading to the privileging of particular contrast classes, descriptive groupings, and levels of explanation, and calling attention to specific observable facts, all of which might have been otherwise for another observer or community of observers.[8] Thus it is possible to maintain a realistic perspective about the targets of measurement – objects, systems, and their attributes – while acknowledging that knowledge is constructed by humans, and can be constructed in multiple ways.

However, while this somewhat softer formulation of realism may accord well with the intuition of many scholars, it leaves open the precise interpretation of many aspects of measurement claims made in the psychological sciences, such as the meaning of comparisons of quantities (e.g., the claim that one student learned "twice as much" as another). In the more formal literature on realism in psychological measurement, Michell (e.g., [48]) has argued that the 'classical' understanding of the concept of measurement – the estimation of ratios of quantities – entails realism about entities, attributes, and numbers (instantiated as ratios between magnitudes). A key consequence of this 'classical' understanding as presented by Michell is that the world is divided into two mutually exclusive natural kinds, namely, what is measurable (i.e., quantities) and what is not (i.e., everything else). Identifying whether an attribute of interest belongs to the former or the later category is what Michell [45] calls the "scientific task of quantification" (p. 75), which is necessarily prior to any attempt at measurement. How exactly one may test this hypothesis is still a matter of some debate (see, e.g., [47,80,49]), though if by whatever method this attribute is determined not to be quantitative, then it is not measurable, and we must instead seek alternative methods for gaining knowledge about what students know and can do in this domain. If the attribute is determined to be quantitative, the "instrumental" task is to then find methods for estimating persons' magnitudes of the attribute, relative to some unit of this attribute, which may be set by convention. Successfully making determinations of such magnitudes constitutes successful measurement.

In Michell's view, measurement requires that the measured attribute possess quantitative structure, in the sense of conforming to Hölder's axioms of quantity. However, it is not clear that there is any *a priori* reason to accept that the term "measurement" should only be used in situations in which an attribute possesses quantitative structure [39]. Other sources do not impose this restriction: for example, in the International Vocabulary of Metrology[9] it is stated that attributes with only ordinal relations are measurable, and some realists (e.g., [9]) use the concept of measurement in reference to attributes with purely nominal differences. Additionally, it is not an uncontroversial assertion that even many paradigmatic physical attributes such as electrical charge possess quantitative structure in the sense given by Hölder (insofar as there exists a lower bound on possible electrical charge, i.e., the "elementary" charge, or the charge of a single electron, in apparent violation of Hölder's second axiom). Finally, on constructive-realist views such as Messick's [42]) or Mislevy's [52], it is not obvious how one would interpret the claim that an attribute *possesses* quantitative structure, given that an attribute label is used to refer to a (potentially very large) collection of more specific facts about an individual, and linguistic devices used for efficient communication are generally not the sorts of things held to be capable of possessing any sort of structure. This said, as previously noted, results of educational and psychological testing procedures are very commonly presented in quantitative terms, without significant qualification, suggesting some continuing confusion regarding the semantics of measurement claims made by psychological scientists.

5. Conclusion

Empiricism is motivated by the intuition that the preferred method of acquiring knowledge is through observation and experience. Pragmatism is motivated by the intuition that theories are made useful via their links with practice. Realism is motivated by the intuition that scientific inquiry seeks to gain knowledge about a natural world. None of these intuitions contradicts the others; however, severe formulations (or, one might argue, misunderstandings) of each of these positions may wind up either saying something false or neglecting to say something important and true by denying the useful intuitions of the other views. In our view, extreme formulations of each position are problematic: (a) strong versions of operationalism and other views associated with logical positivism and behaviourism overreach in denying that attributes exist independently of observations, and more generally in insisting that observations can exhaust the meaning of theoretical concepts; (b) strong forms of pragmatism overreach in denying or ignoring the need for the justification and evaluation of measurement claims embedded in larger claims about test use; and (c) strong

[8] Elsewhere [40], we have discussed how Searle's (e.g., [64]) distinction between ontological and epistemic subjectivity and objectivity, and his recognition of the existence of intentionality-dependent objects and attributes of objects, provides the conceptual vocabulary with which a coherent realist account of the ontology of psychological attributes can be formulated. Briefly, psychological attributes can (a) involve conscious phenomena with subjective first-person ontology, and (b) have conceptual boundaries delineated by contextually and pragmatically-driven frames of reference, rather than being natural kinds (or attributes) in the classic sense. Further, the connection between natural reality and the outcomes of a measurement procedure is not in itself compromised by the fact that we choose to privilege certain contrast classes, levels of explanation, methods of summarisation, and modes of description – that is, that we choose to *model* the world in a particular way. In fact, scientific models (including both substantive models, such as a cognitive-developmental model of learning in a particular domain, and statistical models, such as the Rasch model) serve precisely the purpose of organising experience.

[9] http://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf.

Please cite this article in press as: A. Maul et al., On the philosophical foundations of psychological measurement, Measurement (2015), http://dx.doi.org/10.1016/j.measurement.2015.11.001

versions of realism overreach in denying that knowledge is constructed by humans, and that it can be accurately and usefully constructed in multiple ways depending on the observer's frame of reference and practical concerns.

This stated, conceptual pluralism is not the same as pure relativism: responsible science requires awareness and acknowledgement of the role that conceptual frameworks, methodological approaches, and statistical models play in shaping investigations, and requires explication and empirical investigation of the hypothesised connections between the psychological realities under investigation and the outcomes of purported measurement procedures. Such awareness and acknowledgement is only possible to the extent to which claims are made explicit, and explication of claims requires a coherent semantics with which claims can be formulated. Although we have not attempted to articulate such a set of semantics in this paper, we hope that future work in this direction may benefit from our elucidation of the philosophical foundations of psychological measurement.

References

[1] E.W. Adams, On the nature and purpose of measurement, Synthese 16 (2) (1966) 125–169.
[2] AERA, APA, & NCME, Standards for educational and psychological testing. American Psychological Association, Inc., Washington, DC, 2014.
[3] M.H. Bickhard, The tragedy of operationalism, Theory Psychol. 11 (1) (2001) 35–44.
[4] R.J. Bernstein, The Pragmatic Turn, Polity, 2010.
[5] S. Blackburn, Truth: A Guide, Oxford University Press, 2005.
[6] Jim Bogen, Theory and Observation in Science, The Stanford Encyclopedia of Philosophy (Spring 2013 Edition), in: Edward N. Zalta (Ed.), URL = <http://plato.stanford.edu/archives/spr2013/entries/science-theory-observation/>.
[7] E.G. Boring, Intelligence as the tests test it, New Republic (1923) 35–37.
[8] D. Borsboom, Measuring the Mind: Conceptual Issues in Contemporary Psychometrics, Cambridge University Press, 2005.
[9] D. Borsboom, Latent variable theory, Measurement: Interdisciplinary Issues Perspect. 6 (2008) 25–53.
[10] D. Borsboom, Educational measurement, fourth ed.: book review, Struct. Equation Model. 16 (2009) 702–711.
[11] D. Borsboom, A.O.J. Cramer, R.A. Kievit, A. Zand Scholten, S. Franic, The end of construct validity, in: R.W. Lissitz (Ed.), The Concept of Validity, Information Age Publishing, 2009, pp. 135–170.
[12] D. Borsboom, G.J. Mellenbergh, J. Van Heerden, The theoretical status of latent variables, Psychol. Rev. 110 (2003) 203–219.
[13] D. Borsboom, G.J. Mellenbergh, J. Van Heerden, The concept of validity, Psychol. Rev. 111 (2004) 1061–1071.
[14] D. Borsboom, A. Zand Scholten, The Rasch model and conjoint measurement theory from the perspective of psychometrics, Theory Psychol. 18 (2008) 111–117.
[15] P.W. Bridgman, The Logic of Modern Physics, Macmillan, New York, 1927.
[16] N.R. Campbell, Physics: the Elements, Cambridge University Press, 1920.
[17] H. Chang, Operationalism", *The Stanford Encyclopedia of Philosophy* (Fall 2009 Edition), Edward N. Zalta (Ed.), URL = <http://plato.stanford.edu/archives/fall2009/entries/operationalism/>.
[18] N. Chomsky, Review of verbal behavior by B.F. Skinner, Language 35 (1959) 26–58.
[19] N. Cliff, Abstract measurement theory and the revolution that never happened, Psychol. Sci. (1992).
[20] L.J. Cronbach, G.C. Gleser, H. Nanda, N. Rajaratnam, The Dependability of Behavioral Measurements, Wiley, New York, 1972.
[21] L.J. Cronbach, P.E. Meehl, Construct validity in psychological tests, Psychol. Bulletin 52 (4) (1955) 281.
[22] E.E. Cureton, Validity, in: E.F. Lindquist (Ed.), Educational Measurement, American Council on Education, Washington, DC, 1951, pp. 621–694.
[23] O.D. Duncan, Notes on Social Measurement Historical and Critical, vol. 1, Russell Sage Foundation, 1984.
[24] L. Finkelstein, Widely, strongly and weakly defined measurement, Measurement 34 (2003) 39–48.
[25] L. Finkelstein, Problems of measurement in soft systems, Measurement 38 (2005) 267–274.
[26] C.D. Green, Of immortal mythological beasts: operationism in psychology, Theory Psychol. 2 (1992) 291–320.
[27] C.D. Green, Operationalism again: what did Bridgman say?, Theory Psychol 11 (2001) 45–51.
[28] S. Humphry, Understanding measurement in light of its origins, Front. Psychol. 4 (2013) 113, http://dx.doi.org/10.3389/fpsyg.2013.00113.
[29] W. James, Pragmatism: a New Name for Some old Ways of Thinking, Harvard University Press, Cambridge, MA, 1907.
[30] JGCM, International Vocabulary of Metrology (VIM) – Basic and General Concepts and Associated Terms (2008 edition with minor corrections), Joint Committee for Guides in Metrology, <http://www.bipm.org/en/publications/guides/vim.html>, 2012.
[31] M.T. Kane, Validation, in: R.L. Brennan (Ed.), Educational Measurement, fourth ed., Greenwood Publishing Group, Santa Barbara, 2006, pp. 17–64.
[32] D. Khlentzos, Challenges to Metaphysical Realism, The Stanford Encyclopedia of Philosophy (Spring 2011 Edition), Edward N. Zalta (Ed.), URL = <http://plato.stanford.edu/archives/spr2011/entries/realism-sem-challenge/>, 2011.
[33] D.H. Krantz, R.D. Luce, P. Suppes, A. Tversky, Foundations of Measurement, Additive and Polynomial Representations, vol. 1, Academic Press, New York, 1971.
[34] A. Kyngdon, The Rasch model from the perspective of the representational theory of measurement, Theory Psychol. 18 (2008) 89–109.
[35] A. Kyngdon, Plausible measurement analogies to some psychometric models of test performance, British J. Math. Statistical Psychol. 64 (2011) 478–497.
[36] A. Kyngdon, Descriptive theories of behaviour may allow for the scientific measurement of psychological attributes, Theory Psychol. 23 (2013) 227–250.
[37] F.M. Lord, M.R. Novick, A. Birnbaum, Statistical Theories of Mental Test Scores, vol. 47, Addison-Wesley, Reading, MA, 1968.
[38] L. Mari, P. Carbone, D. Petri, Measurement fundamentals: a pragmatic view, IEEE Trans. Instrument. Measurement 61 (8) (2012) 2107–2115, http://dx.doi.org/10.1109/TIM.2012.2193693.
[39] L. Mari, A. Maul, D. Torres Irribarra, M. Wilson, Quantification is neither necessary nor sufficient for measurement, J. Phys.: Conf. Series 459 (2013), http://dx.doi.org/10.1088/1742-6596/459/1/012007.
[40] A. Maul, On the ontology of psychological attributes, Theory Psychol. 23 (2013) 752–769.
[41] J. McGrane, Stevens' forgotten crossroads: the divergent measurement traditions in the physical and psychological sciences from the mid-twentieth century, Front. Psychol. 6 (2015) 431, http://dx.doi.org/10.3389/fpsyg.2015.00431.
[42] S. Messick, Validity, in: R.L. Linn (Ed.), Educational Measurement, third ed., American Council on Education/Macmillan, New York, 1989, pp. 13–103.
[43] J. Michell, An introduction to the Logic of Psychological Measurement, Lawrence Erlbaum Associates, 1990.
[44] J. Michell, Quantitative science and the definition of measurement in psychology, British J. Psychol. 88 (1997) 355–383.
[45] J. Michell, Measurement in Psychology: a Critical History of a Methodological Concept, Cambridge University Press Cambridge, England, 1999.
[46] J. Michell, Normal science, pathological science and psychometrics, Theory Psychol. 10 (5) (2000) 639.
[47] J. Michell, Item response models, pathological science and the shape of error: reply to Borsboom and Mellenbergh, Theory Psychol. 14 (1) (2004) 121.
[48] J. Michell, The logic of measurement: a realist overview, Measurement 38 (4) (2005) 285–294.
[49] J. Michell, Is psychometrics pathological science?, Measurement: Interdisciplinary Res Perspective 6 (1) (2008) 7–24.
[50] J. Michell, Qualitative research meets the ghost of Pythagoras, Theory Psychol. 21 (2011) 241–259.
[51] G.A. Miller, The cognitive revolution: a historical perspective, Trends Cogn. Sci. 7 (3) (2003) 141–144.

Please cite this article in press as: A. Maul et al., On the philosophical foundations of psychological measurement, Measurement (2015), http://dx.doi.org/10.1016/j.measurement.2015.11.001

[52] R. Mislevy, Validity from the perspective of model-based reasoning, in: R.W. Lissitz (Ed.), The Concept of Validity: Revisions, New Directions and Applications, Information Age Publishing, Charlotte, NC, 2009, pp. 83–108.
[53] P.C.M. Molenaar, A. Von Eye, On the arbitrary nature of latent variables, in: A. von Eye, C.C. Clegg (Eds.), Latent Variables Analysis, Sage, Thousand Oaks, 1994.
[54] L. Narens, R. Luce, Measurement: the theory of numerical assignments, Psychol. Bulletin 99 (2) (1986) 166–180.
[55] R.E. Nisbett, J. Aronson, C. Blair, W. Dickens, J. Flynn, D.F. Halpern, E. Turkheimer, Intelligence: new findings and theoretical developments, Am. Psychol. 67 (2012) 130–159.
[56] C.S. Pierce, The essential peirce, in: N. Houser, C. Kloesel, & the Peirce Edition Project (Eds.), Bloomington: Indiana University Press, 1998.
[57] H. Putnam, Realism and Reason (volume 3 of Philosophical Papers), Cambridge University Press, Cambridge, 1985.
[58] H. Putnam, Realism with a Human Face, Harvard University Press, Cambridge, MA, 1990.
[59] H. Putnam, The Dewey lectures, J. Philos. 91 (1994) 445–517.
[60] H. Putnam, The Threefold Cord: Mind, Body and World, Columbia University Press, New York City, 2000.
[61] R. Rorty, Philosophy and Social Hope, Penguin, 1999.
[62] R. Rorty, M. Williams, D. Bromwich, Philosophy and the Mirror of Nature, vol. 401, Princeton University Press, Princeton, NJ, 1980.
[63] U. Saint-Mont, What measurement is all about, Theory Psychol. 22 (2012) 467–485.
[64] J. Searle, The Rediscovery of the Mind, MIT Press, 1992.
[65] P. Schönemann, Measurement: the reasonable ineffectiveness of mathematics in the social sciences, in: I. Borg, P. Mohler (Eds.), Trends and Perspectives in Empirical Social Research, Walter de Gruyter, Berlin, Germany, 1994, pp. 149–160.
[66] L.A. Shepard, The centrality of test use and consequences for test validity, Educ. Measurement: Issues Practice 16 (2) (1997) 5–8.
[67] K. Sijtsma, Psychological measurement between physics and statistics, Theory Psychol. 22 (2012) 786–809.
[68] B.F. Skinner, The operational analysis of psychological terms, Psychol. Rev. 52 (5) (1945) 270.
[69] B.F. Skinner, Beyond Freedom and Dignity, Knopf, New York, 1971.
[70] S.S. Stevens, On the theory of scales of measurement, Science 103 (2684) (1946) 677–680.
[71] S.S. Stevens, Psychophysics, Transaction Publishers, 1975.
[72] P. Suppes, J.L. Zinnes, Basic measurement theory, Handbook Math. Psychol. 1 (1–76) (1963).
[73] E. Tal, Old and new problems in the philosophy of measurement, Philosophy Compass 8 (2013) 1159–1173.
[74] E. Tal, Measurement in Science The Stanford Encyclopedia of Philosophy (Summer 2015 Edition), in: Edward N. Zalta (Ed.), URL = <http://plato.stanford.edu/archives/sum2015/entries/measurement-science/>, 2015.
[75] D. Torres Irribarra, A Pragmatic Perspective on Measurement, Paper Presented at the 2014 International Meeting of the Psychometric Society, Madison, Wisconsin, July 2014.
[76] S.E. Toulmin, The Uses of Argument, Cambridge University Press, 1958.
[77] G. Trendler, Measurement theory, psychology and the revolution that cannot happen, Theory Psychol. 19 (5) (2009) 579–599.
[78] G. Trendler, Measurement in psychology: a case of ignoramus et ignorabimus? A rejoinder, Theory Psychol. (2013), http://dx.doi.org/10.1177/0959354313490451.
[79] D. Borsboom, The attack of the psychometricians, Psychometrika 71 (3) (2006) 425–440.
[80] D. Borsboom, G. Mellenbergh, Why psychometrics is not pathological: a comment on Michell, Theory Psychol. 14 (2004) 105–120.
[81] P.W. Bridgman, P.W. Bridgman's "The Logic of Modern Physics" after thirty years, Daedalus 88 (1959) 518–526.
[82] D.T. Campbell, D.W. Fiske, Convergent and discriminant validation by the multitrait-multimethod matrix, Psychol. Bull. 56 (2) (1959) 81.
[83] M. Devitt, Realism and Truth, Blackwell, Oxford, 1991.
[84] B.D. Haig, Investigating the Psychological World: Scientific Method in the Behavioral Sciences, MIT Press, 2014.
[85] T.L. Kelley, Interpretation of Educational Measurements, Macmillan, New York, 1927.
[86] T.H. Leahy, The myth of operationism, J. Mind Behav. 1 (2) (1980) 127–144.
[87] K.A. Markus, D. Borsboom, Frontiers of Test Validity Theory: Measurement, Causation and Meaning, Routledge, 2013.
[88] R. Perline, B.D. Wright, H. Wainer, The Rasch model as additive conjoint measurement, Appl. Psychol. Meas. 3 (1979) 237–255.
[89] D. Sherry, Thermoscopes, thermometers, and the foundations of measurement, Studi. Hist. Philos. Sci. 42 (2011) 509–524.
[90] J.D. Trout, Measuring the Intentional World: Realism, Naturalism, and Quantitative Methods in the Behavioral Sciences, Oxford University Press, New York, 1998.
[91] B.D. Zumbo, Validity as contextualized and pragmatic explanation, and its implications for validation practice, in: R.W. Lissitz (Ed.), The Concept of Validity: Revisions, New Directions and Applications, 2009, pp. 65–82.
[92] J. Michell, Rejoinder, Meas.: Interdiscipl. Res. Perspect. 6 (1) (2008) 125–133.

Please cite this article in press as: A. Maul et al., On the philosophical foundations of psychological measurement, Measurement (2015), http://dx.doi.org/10.1016/j.measurement.2015.11.001