Beyond psychometrics

Measurement, non-quantitative structure, and applied numerics

Paul Barrett

Mariner7 Ltd, Manukau, Auckland, New Zealand, Department of Psychology, University of Auckland, Auckland, New Zealand, Department of Psychology, University of Canterbury, Christchurch, New Zealand, and Department of Clinical Psychology, University of Liverpool, Liverpool, UK

Keywords *Psychometric tests, Numerical methods, Quantitative modelling*

Abstract *A statement from Michell (Michell, J., "Normal science, pathological science, and psychometrics",* Theory and Psychology, *Vol. 10 No. 5, 2000, pp. 639-67), "psychometrics is a pathology of science", is contrasted with conventional definitions provided by leading texts. The key to understanding why Michell has made such a statement is bound up in the definition of measurement that characterises quantification of variables within the natural sciences. By describing the key features of quantitative measurement, and contrasting these with current psychometric practice, it is argued that Michell is correct in his assertion. Three avenues of investigation would seem to follow from this position, each of which, it is suggested, will gradually replace current psychometric test theory, principles, and properties. The first attempts to construct variables that can be demonstrated empirically to possess a quantitative structure. The second proceeds on the basis of using qualitative (non-quantitatively structured) variable structures and procedures. The third, applied numerics, is an applied methodology whose sole aim is pragmatic utility; it is similar in some respects to current psychometric procedures except that "test theory" can be discarded in favour of simpler tests of observational reliability and validity. Examples are presented of what future practice may look like in each of these areas. It is to be hoped that psychometrics begins to concern itself more with the logic of its measurement, rather than the ever-increasing complexity of its numerical and statistical operations.*

Introduction

Consider the following definitions of psychometrics from a sample of current text: Kline (2000, p. 1) defines psychometrics as "[p]sychometrics refers to all those aspects of psychology which are concerned with psychological testing, both the methods of testing and the substantive findings". Cronbach (1990, p. 34) refers to psychometrics as "[p]sychometric testing sums up performance in numbers. Its ideal is expressed in two famous old pronouncements: If a thing exists, it exists in some amount, and, if it exists in some amount, it can be measured". Suen (1990, p. 4) defines it as "[t]he science of developing educational and psychological tests and measurement procedures has become highly sophisticated and has developed into such a large body of knowledge that it is considered a scientific discipline of enquiry in its own right. This discipline is referred to as *psychometrics*". McDonald (1999, p. 1) refers to psychometric theory as "[t]est theory is an abbreviated expression for *theory of*

Journal of Managerial Psychology
Vol. 18 No. 5, 2003
pp. 421-439

0268-3946
DOI 10.1108/02683940310484026

psychological tests and measurements, which in turn can be abbreviated back to *psychometric theory* (psychological measurement)". Finally, Miles (2001, p. 62) defines psychometrics as "… the branch of psychology concerned with studying and using measurement techniques".

The latter definitions would appear to indicate that psychometrics is totally concordant with the goals of measurement and science, yet Michell (2000, p. 639) states "[i]t is concluded that psychometrics is a pathology of science …" and Michell (2001, p. 211) says "… the way in which psychometrics is currently, typically taught actually subverts the scientific method". What are we to make of the contrast between these two positions? There are four critical points of understanding to be addressed.

1. Quantitative measurement

Michell (2001, p. 212) states:

> Measurement, as a scientific method, is a way of finding out (more or less reliably) what level of an attribute is possessed by the object or objects under investigation. However, because measurement is the assessment of the of a level of an attribute via its numerical relation (ratio) to another level of the same attribute (the unit selected), and because only quantitative attributes sustain ratios of this sort, measurement applies only to quantitative attributes. Psychometrics concerns the measurement of psychological attributes using the range of procedures collectively known as psychological tests. As a precondition of psychometric measurement, these attributes must be quantitative.

This definition is absolutely clear, technical, and precise. It introduces the concept of a "quantitative variable" (one whose values are defined by a set of ordinal and additive relations). Further, such variables require a unit of measurement to be explicitly identified, such that magnitudes of a variable may be expressed relative to that unit. Thus, as stated in the second passage, "measurement applies only to quantitative variables".

2. Quantitatively structured variables

A variable is anything relative to which objects may vary. For example, weight is a variable, different objects can have different weights, but each object can only possess one such weight at any point in time. A quantitative variable satisfies certain conditions of ordinal and additive structure. For example, weight is a quantity because weights are ordered according to their magnitude, and each specific weight is constituted additively of other specified weights. Likewise lengths. More formally, let X, Y, and Z be any three values of a variable, Q. Then Q is ordinal if and only if:

(1) $X \geq Y$ and $Y \geq Z$ then $X \geq Z$ (transitivity);

(2) $X \geq Y$ and $Y \geq X$ then $X = Y$ (antisymmetry);

(3) either $X \geq Y$ or $Y \geq X$ (strong connexity).

A relation possessing these three properties is called a simple order, so Q is ordinal if and only if $\geq$ is a simple order on its values. All quantitative variables are simply ordered by $\geq$, but not every ordinal variable is quantitative, for quantity involves more than order. It involves additivity.

Additivity is a ternary relation (involving three values), symbolized as "$X + Y = Z$". Let Q be any ordinal variable such that for any of its values X, Y, and Z:

(4) $X + (Y + Z) = (X + Y) + Z$ (associativity);

(5) $X + Y = Y + X$ (commutativity);

(6) $X \geq Y$ if and only if $X + Y \geq Y + Z$ (monotonicity);

(7) if $X > Y$ then there exists a value of Z such that $X = Y + Z$ (solvability);

(8) $X + Y > X$ (positivity);

(9) there exists a natural number n such that $nX \geq Y$ (where $1X = X$ and $(n + 1)X = nX + X$) (the Archimedian condition).

In such a case the ternary relation involved is additive and Q is a quantitative variable.

These nine conditions were stated by J.S. Mill in 1843 in *A System of Logic* (see Robson, 1973), and later by Hölder (1901) (translated in Michell and Ernst, 1996, 1997) within his exposition of the axioms of quantity. However, as Michell (1999) points out, the influence of Euclid's theory of magnitudes is present throughout the historical development of the physical sciences, and especially within Newton's *Principia* of 1728. In short, these are the bases for the kind of quantitative measurement that has evolved within the natural sciences.

3. Numbers and their status

The question that now arises is that of the status of numbers. If we treat numbers as an abstract system of symbols, that can be assigned as and how a scientist decides they should be used to represent objects within an empirical relational system, then we have representationalism.

A representational theory of measurement in its broadest sense, states that measurement requires defining how an empirical relational system may be conjoined with a number system in order to permit an individual to describe "quantities" of empirical entities using these numbers. An empirical relational system like weight possesses an ordered structure with the relations defined earlier. For example, if a class of objects that possess the attribute weight can be compared to one another with a relation such as "being at least as heavy as", then the weights standing in this relation to one another are said to constitute a relational system. In essence, a comparison operation is required to take place between all objects in this system in order to determine whether the relation holds for any two such objects, and to observe whether the properties of the relations expressed in abstract form earlier can also be observed using the objects that are said to possess weight. A numerical relational system is one in which the entities involved are numbers, and the relations between them are numerical relations. An example of a numerical relation is the set of all positive integers less than say 1,000, with the relation of "being at least as great as". Each number can be compared to another and a determination made as to whether the relation holds for that pair. We can also apply such relations to real numbers, and observe the properties of the same relations but now using continuous quantities rather than discrete values. So, in the case of weight, the numerical representation of weight is achieved by matching numbers to objects so that the order of weights of objects is reflected in the order (magnitude) of the numbers.

It was Stevens (1951, p. 23) who proposed this kind of measurement theory as describing measurement within psychology:

> ... in dealing with the aspects of objects we can invoke empirical operations for determining equality (the basis for classifying things), for rank ordering, and for determining when differences and ratios between the aspects of objects are equal ... This isomorphism between the formal system and the empirical operations performed with material things justifies the use of the formal system as a model to stand for aspects of the empirical world.

Thus, any numerical modelling of an empirical system constitutes measurement. Stevens (1959) stated perhaps the more familiar exposition of this statement as measurement as the assignment of numbers to objects by rule and that "provided a consistent rule is followed, some form of measurement is achieved" (Stevens, 1959, p. 19).

This seems a reasonable statement on the surface, but, it is flawed. In order to assign a numerical system to an empirical relational system, it was required that the empirical relations could first be identified without necessarily assigning numbers to objects within the system. It was a prior requirement that whether or not an empirical relation possesses certain properties was a matter for empirical, scientific investigation. As Michell (1999, p. 168) states:

> Simply to presume that a consistent rule for assigning numerals to objects represents an empirical relation possessing such properties is not discover that it does; it is the opposite.

For, what Stevens was really saying is that it is not the independently existing features of objects (the properties or relations of objects) that are represented in measurement, but that the numerical relations imposed by an investigator in fact determine the empirical relations between objects.

If one considers the real number relational system defined within the continuous theory of measurement to be an empirical fact (Michell, 1994) in its own right, and that the conjoining of this system to an empirical relational system (also considered to be a putative or actual fact by an investigator) is an empirical hypothesis rather than an assertion by an investigator, then the representationalism espoused by Stevens and psychologists since 1951 is seen to be an impediment to any form of scientific investigation, and not as Stevens saw it, a different kind of measurement construction that was applicable especially to the social science. To complete the picture, a definition of the process of quantification is perhaps the best way of summarising the content of the three points above.

4. The process of quantification

Michell (1999, p. 75) states:

> Because measurement involves a commitment to the existence of quantitative attributes, quantification entails an empirical issue: is the attribute involved really quantitative or not? If it is, then quantification can sensibly proceed. If it is not, then attempts at quantification are misguided … The logically prior task in this enterprise is that of addressing this empirical issue. I call it the scientific task of quantification.

It is to be hoped that the reader can now see why Michell (2000) calls psychometrics a pathology of science. It assigns numbers to attributes without ever considering whether those attributes can sustain the operations represented within the empirical numeric relation system so imposed. To assume that the manipulation of numerals that are imposed from an independent relation system can somehow discover facts about other empirical objects, constructs, or events is "delusional", just as Michell (1997) stated. But why have psychologists been so adamant in equating measurement with psychological science?

The Pythagorean or "measurement imperative"

The idea that for anything to be considered "scientific" it must somehow involve quantitative measurement, has evolved from Pythagoras (approximately during the sixth century BC). His philosophy stated that nature and reality was revealed through mathematics and numerical principles. These numerical principles were proposed as explaining psychological as well as physical phenomena. Given that mathematics might provide the principles by which all phenomena might be understood, and given it can be considered the science of structure (Parsons, 1990; Resnick, 1997), then it is reasonable to assume that mathematics could indeed be the means by which nature and reality might be understood. This was the driving philosophy behind the Scientific Revolution in the seventeenth century. With the success of quantitative physics in the nineteenth century, came an almost absolute certainty that what could not be measured was of no substantive scientific import. The Kelvin dictum was born during that century (Thomson, 1891, pp. 80-1):

> I often say that when you can measure what you are speaking about and express it in numbers, you know something about it; but when you cannot measure it, when you cannot

> express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge but you have scarcely in your thoughts advanced to the stage of science, whatever the matter may be.

This was the dictum that threatened the fledgling science of psychology at its very beginning. If it was to be considered a science by others, it had to make measurement in the manner of the physical sciences. This was reinforced by Thorndike (1918):

> Whatever exists at all exists in some amount. To know it thoroughly involves knowing its quality as well as its quantity.

During this period in psychology, "practicalism" also became the modus operandi, along with the Pythagorean view. This is illustrated by a quotation from Kelley (1929, p. 86), summing up the position that intelligence is a measurable variable:

> Our mental tests measure something, we may or may not care what, but it is something which it is to our advantage to measure, for it augments our knowledge of what people can be counted upon to do in the future. The measuring device as a measure of something that it is desirable to measure comes first, and what it is a measure of comes second.

The problem with the original and neo-Pythagorean views is that they assume that all structures, entities, and phenomena can be described by the mathematics of quantity, using quantitatively structured variables. That much of the natural sciences could be described in this manner was taken as the signal that psychological constructs could be similarly measured, albeit with some initial difficulty. The original philosophy of Pythagoras had been distorted through the seventeenth through nineteenth centuries into a kind of measurement imperative. If a discipline could not demonstrate measurement of its constructs and variables, then it could not be considered a science. But science is a method or process for the investigation of phenomena. It does not require that the variables within its domain of enquiry be quantitatively structured. Quantitative science does demand such properties of its variables. Therein lies the simple yet fundamental distinction between a quantitative science and a non-quantitative science.

Psychological "measurement" as "something different"

It is quite possible to retain recognition of the axioms of quantity, yet still proceed to argue that psychology is a "special science" that may require a different approach to understanding causality than the physical sciences (via some version of non-linear complex or non-quantitative methods). Even in Quantum mechanics (which is often used as an exemplar for "a different kind of measurement") where uncertainty prevails in any measurement of the state of a system under a set of given conditions, the constituent system variables are themselves measurable as quantitative variables. For example, quantum computation using Qubits relies upon accurate quantitative measurement of absolute temperature in order to control coherence, as well as the quantitatively measurable components of electrical activity (Vion *et al.*, 2002). In short, it is not the measurement principles that change to suit relevant explanatory theory, but the very structure of the variables and the subsequent relations between them.

Those for example who use multivariate statistical techniques such as regression analysis, factor analysis, structural equation modelling, hierarchical multilevel analysis, etc. are applying arithmetic operations that rely upon the properties of ordinal and additively structure variables. The problem is not one of "permissible statistics" or that one cannot produce numerical results from such techniques, but, the status of any conclusions drawn remains in doubt while the quantitative structure of the variables so manipulated remains untested.

However, even accepting the above might well be true, psychologists will then proceed to quote the doctrine of practicalism. The argument goes something like "regardless of whatever it is that psychologists do when they claim to be measuring something, in many areas a substantive body of knowledge has been crafted and created using the tools and techniques of quantitative science". Therefore, it is concluded that because of these practical and useful results which have real-world implications, the measurement issue is really a non-issue or of only minor importance. This reflects the approach taken by Thorndike, espoused as early as 1904, that test scores may not reflect some quantitatively structured variable such as "ability", but they can be rank ordered, and by expressing the relative positions amongst the score range using operations such as re-expressing scores as standardised values, measurement with something of the accuracy and precision of physical variables could be achieved:

> Measurement by relative position in a series gives as true, and may give as exact, a means of measurement as that by units of amount (Thorndike, 1904, p. 19).

However, such "measurement" may simply be just a monotonic transformation of observed test scores. The problem remains with what the test scores are actually measures of; that is, what is the empirical relation-order structure of the variable, which is used to explain the occurrence of the test scores? A quantitatively structured variable possesses a unit of quantity against which all other amounts of a variable are to be compared. This unit is required to be made explicit within any quantitative measurement operation.

The question of whether it matters is of immediate concern to scientists who wish to understand how the human mind works and to provide causal explanations for behaviours; for it is the role of a scientist to seek explanations for phenomena, not merely to provide numerical indices that have some immediate practical value or that provide some illusion of "explanatory coherence". It is the thesis of this paper that given the comments above, most of current psychometrics can no longer continue to be viewed as a "series of methods, theory, and techniques for producing measurement of psychological constructs". It may or may not be producing such measurement; for the measurability hypothesis for a single variable remains untested and therefore retains the status of an assumption.

The way forward

From the above exposition, it is suggested that three avenues are now open to an investigator. First, there is an approach that espouses measurement in accordance with the axioms and content of the critical points (1)-(4) above. The second is one that adopts a philosophical view that psychological attributes are non-quantitative, and hence seeks to construct a body of knowledge based solely upon partial order structured variables (ordinal relations). Third, there is the avenue that I call "applied numerics". This approach encompasses the kind of "measurement" of magnitudes of psychological variables using classical test theory, two- and three-parameter item response theory, and the manipulation of test scores and variable magnitudes that use linear additive operations (e.g. the techniques that use means, variances, and covariances as the components of analysis).

Avenue 1: measurement

The problem that faces psychology is that the variables that are of most interest to investigators are latent or unobservable. That is, they do not exist as physical objects or material, which can be manipulated in order to determine the empirical relations that may hold between amounts of an object (like the length of wooden rods for example). Psychological variables such as intelligence, motivation, personality, self-esteem, anger, religiosity, beliefs etc. do not "exist" except as inferred constructs. Within physics, a similar problem could be perceived with "derived" measures such as "density". Density is not a physical object with observable units that can be physically concatenated or manipulated. It is derived from the operation of two other physical measures which can be manipulated, mass and volume. The operation between these two "extensive" variables is that of division – taking the ratio of mass to volume yields a value for the variable density. For each substance, the ratio of mass to volume is a constant. What was intriguing to some was how it could be proven that the combination of two variables could produce a third whose values were themselves ordinal and additively structured in the manner of a quantitative variable.

Luce and Tukey (1964) published the axioms of conjoint measurement, the necessary set of conditions that if met by combining values of any set of three variables, would provide empirical proof of the additive structure of all three variables. While this might have been of minor importance to psychologists had it been confined to dealing with extensive (already quantitatively structured) measures such as mass and volume, it was not. Luce and Tukey showed that even if all three variables possessed values that were simply ordered (ordinal relations), then by combining these values in order to test for three special conditions, and meeting the conditions as specified, then all three variables could be considered as possessing quantitative structure. Krantz *et al.* (1971) have since provided the complete set of formal proofs for the conjoint measurement axioms. Michell (1990, chapter 4) provides a detailed yet understandable exposition of the axioms and worked example of this procedure.

Examples of conjoint measurement using explicit tests of the three conjoint axioms within psychology are rare – however, an interesting one is that provided in Stankov and Cregan (1993) that examines the hypothesis that intelligence (as proposed to be measured by the number of items correct on a Letter Series task) could be considered a quantitative variable, measured conjointly by working memory capacity and motivation. They suggest that Intelligence = f(M, WM); that is, Intelligence is some mathematical function of motivation and working memory.

Of critical importance is the realisation that the currently fashionable Rasch item response theory is also an empirical instantiation of the conjoint additivity axioms (Perline *et al.*, 1979). That is, the construction of a latent variable using Rasch item analysis is no less than the empirical test of quantitative structure for that latent variable. The significance of this fact for psychological measurement cannot be underestimated, because it allows Psychology to be reconciled with the axioms outlined above. Bond and Fox (2001) provide what is currently the best and most easily understood introduction to Rasch modelling, and demonstrate both the simplicity and desirability of constructing quantitatively structured variables. The Institute for Objective Measurement in Chicago is devoted to the theory, procedures, and methods for the construction of quantitative measures. Its members produce measures within a wide domain of investigation, from medicine, education, sociology, through to psychology.

While the construction of variables that possess quantitative structure is now possible within psychology, a priori meaning instantiation remains critical. As Barrett (2001, 2002) has indicated, measurement without a clear a priori theory about the nature of the variable to be quantified, is of limited scientific value. This is a point also elaborated upon within Kline's (1998) exposition of the foundations of what he called "the new psychometrics". In essence, Kline was noting that substantive knowledge of psychological attributes and constructs was unlikely to ever be achieved if the debate remained locked around such questions as "which model for measurement is best?". Rasch scaling and additive conjoint measurement are the key tools required by scientists trying to establish empirically that a variable of interest possesses a quantitative structure. However, the task for a science is also explaining why such an empirical finding should be so observed. Simply scaling variables without consideration of whether what has been so scaled is substantively meaningful is a recipe for nonsense, as exemplified by Wood's (1978) demonstration of an almost perfect Rasch scaled latent variable of "coin-tossing" ability.

What the above shows is that it is possible for psychologists to construct and make measurement that accords with the axioms of quantity, in the same way as physical scientists construct and make measurement. It is clear from already existing empirical work that many psychological variables do not possess a quantitative structure, but as Bond and Fox (2001) illustrate, as well as in the many published Rasch scales, some considerable number do. Thus, this is an avenue that psychology may take, with some positive signs already that it is possible to maintain concordance with measurement. However, as Barrett (2002) noted with the variable "g" (the technical definition of the common-sense term "intelligence"), it is also possible to open up completely new domains of research that might potentially yield some much-needed harmonisation of construct understanding and measurement in psychology. This magnitude of challenge and research breadth awaits those who choose this investigatory path.

Avenue 2: non-quantitative variable structures

As Michell (2001) points out, there is no pre-ordained necessity for variables within psychology to possess a quantitative structure. Psychology may remain a science yet deal with both quantitative and qualitative (non-quantitative) variables. Quantity is not synonymous with mathematics. If mathematics is considered as the science of abstract structure then it is obvious that not all structures studied using mathematics are quantitative. For example, the structure of communication and social networks, graphs, language grammars, therapeutic interactions, automata networks, etc. are essentially non-quantitative. The study of them may remain scientific, in that the method of investigation and critical reasoning is applied in accordance with scientific principles, but the variables are a mixture of the quantitative and non-quantitative. A quantitative science is one that relies upon quantitatively structured variables for its measurement. A non-quantitative science relies upon variables that are mainly non-quantitative, using order relations, probabilities of occurrence of discrete behaviours, and structural analysis of data to provide explanatory coherence for its theories.

Perhaps the most obvious psychological example of non-quantitative scientific research is that stemming from Guttman's work with facet theory and the analysis of data structures. Guttman (1971) is an excellent exposition. An entire school of psychology has arisen in Israel, founded on the principles of Guttman's analysis of data structures, rather than quantitatively measured variables (Shye, 1978, 1988). Essentially, this form of analysis uses both nominal (classificatory) and ordinal relations between amounts of any variable. These amounts, generally represented by ranks in the case of ordinal data, are the components of analysis. However, rather than concentrate on producing quantitative measures for variables, and relating these through additive operations, the non-quantitative approach looks for particular kinds of order within data, generally mapping these ordered "sets" in a Euclidean space. However, instead of relying upon the additive units implied in such a space, what is important to this kind of work is the regions in which certain order relations hold for certain variables, and not others. In order to assist the theory construction process, which cannot now rely upon quantity defined by order and additive relations, Guttman introduced facet theory. This allowed a researcher to conceive of theoretically important concepts in terms of facets of structure, which, along with the concept of a mapping sentence (as a means of expressing theoretically important statements in a formal grammar akin to set theory) allowed the computational methods for discovering structure (for example multiple and partial order scalogram analysis, smallest space analysis) to be used as empirical tests of these formally proposed relational structures. Wilson (1995) and Donald (1995) provide introductions to this area of research, while Canter (1983, 1985) provides a thoroughgoing exposition of facet theory. Facet theory has proven to be an extremely versatile and powerful means of relating psychological theory to empirical analysis of data structures.

Another approach to dealing with structure in data is that based upon cellular automata and the science of complex structures and evolved systems (Coveney and Highfield, 1995; Holland, 1998; Wolfram, 1994, 2002). This approach to understanding how complex systems evolve is based upon both mathematical and non-mathematical principles. An evolved system might well begin with a few simple rules that may be defined mathematically, but the evolutionary constraints can be qualitatively structured using order and category relations only, such that the system evolves in a highly non-linear fashion (no additive transformations are possible). Further, Wolfram's work with cellular automata showed how complex structures could evolve in data patterns but for which there was no mathematics to explain the formation of such structures (the concept of a cellular automaton was introduced within computational science by Stanislav Ulam (1952). It is an abstract array of 'cells' that are programmed to implement rules en masse. Each cell may function only in terms of its "nearest neighbour", such that its output is influenced only by those cells adjoining it. This kind of work is maintained as a coherent research strategy at the Santa Fe Institute in the USA (www.santafe.edu), much in the way that Shye and Canter maintain institutes in their respective countries (Israel and the UK) for their non-metric approaches. That these investigatory methods are not even known about in many psychology departments is testament again to the quantitative imperative that pervades current psychological thinking.

Avenue 3: applied numerics

I have introduced this terminology to stand for those classes of mathematical and statistical analyses that rely upon variables possessing ordinal and additive structure, using arithmetic operations that rely upon such properties, yet the hypothesis that these variables might possess these properties of quantity is never tested.

It is within this avenue in which classical and modern two- and three-parameter item response theory are prevalent. Also, the major analytical multivariate techniques of structural equation modelling, regression and exploratory factor analysis may also be found here. While the use of such arithmetic and linear algebraic operations can of course be implemented using the numbers that are said to stand as "measurements", and results so computed, it is the validity of any conclusions drawn that is compromised. For, as stated above, the conclusions drawn do not necessarily follow if the variables used are not quantitatively structured. To have produced test theories such as the classical or two- and three-parameter item response theory models is a testament to the mathematical prowess of the developers of such theory, but the theory is actually disconnected from any scientific study of psychology. Likewise, those who use the very latest developments in psychometrics such as structural equation modelling (SEM), hierarchical multilevel modelling, and latent growth modelling, are just engaging in an approximation exercise of uncertain validity, for little attention is paid to the empirical hypothesis of whether the variables used or introduced as "phantom" latents (Hayduk, 1996) in such models are actually quantitative at all. Instead, these models all rely upon the manipulation of the empirical number system, which is mapped onto an assumed empirical object-entity relational system.

So, as with the many models that invoke concepts of personality and intelligence as causal variables associated with certain phenomena, the knowledge is bound up in the numeric operations applied, rather than in the meaning of what actually constitutes an "intelligence" or "personality" variable. This is a subtle but telling position that becomes apparent when an investigator is asked to explain what it is that the observed test scores are said to be a measurement of, and how such a "cause" comes to possess equal-interval and additive relations between its amounts. This question is no less difficult to answer for a Rasch or additive conjoint measured latent variable. However, in the latter case the investigator can at least be assured that the variable can be shown empirically to possess a quantitative structure. In the case of applied numerics, such as with SEM using assumed quantitative variables, no such knowledge is available. This matters greatly if a theory is proposed that relies for its explanatory coherence upon this structure being a property of some of all of its variables.

While the above constitutes a criticism of psychometrics as a "science" of "psychological measurement, it does not constitute a criticism of it as an approach to the manipulation of numbers that are applied as magnitudes of hypothesised variables, for the purpose of approximating loose theoretical or pragmatic hypotheses. That is, if the process of mapping numbers onto psychological attributes is recognised from the outset as an approximation, with no great regard paid to the scientific value of such an enterprise, then this constitutes an honest approach that has indeed paid many pragmatic dividends. As the history of applied psychometrics has demonstrated, many variables have been constructed and utilised as predictive indicators of practically relevant phenomena (such as job satisfaction, employee wellbeing, personality, IQ), without any explicit theory of the meaning of the variables other than a "common-sense" meaning that is generally applied to assist in their interpretation. Although values for these variables are treated computationally as possessing both ordinal and additive structure, the interpretations of them are invariably made using ordinal relations only. In short, the enterprise is nothing more than an approximation that finds its definition of validity through pragmatic utility. This is not a "scientific" approach, but rather, a pragmatic approach. It is no less important for this, and sometimes the exploration of phenomena in this way does suggest avenues of exploration in a more scientifically-relevant manner. However, such an honest appreciation of the enterprise of applied numerics also opens up new vistas of assessing amounts of psychological variables, for which there need be no particular reliance upon test theoretic constructs such as item universes, item domains, or additive variable assumption statistical models of item or test characteristics. Further, reliability and validity can be simplified into concepts that remain close to observed data (rather than invoking hypothetical "true scores"), with validity defined more by observed pragmatic relevance than some vague notion of "construct validity". In short, the empirical value and stability of the procedures used define their validity, not a test theory that is predicated upon a set of untested assumptions. Necessarily, this limits the knowledge claims that might be made, but this is the price paid by not considering the precise meaning and constituent structure of any variable. That price is traded directly with pragmatic value in applied numerics. Applied examples of this approach can be found in the area of actuarial risk of violence of mentally disordered patients and sex offenders (Quinsey *et al.*, 1998; Doren, 2002) and in the monograph by Swets *et al.* (2000) on making diagnostic decisions using signal detection theory.

In conclusion, I would like to give an example with which I am intimately familiar. Within an organisational psychology area, that of selection and recruitment, an approach that discards conventional test theory in favour of making direct, useful, pragmatic measurement of psychological constructs is the Preference Profile™ technology. What has been achieved here is a form of psychological assessment that does not rely upon questionnaire items as being a sample from some hypothetical universe of items (as in classical test theory), or on a model of uni-dimensional measurement of a latent trait as in item-response theory. Instead, the Preference Profile generates measurement in a manner similar to that which is referred to in clinical psychology as a "repertory grid" procedure, but which is reverse engineered in this case as it provides the fixed, meaningful, dimensions within which an individual will indicate their preferences. This is an entirely computer-enabled graphical method of assessing an individual's job preferences, which are measured using 12 bipolar (opposites) nouns. However, as the design process evolved, it became clear that assessment could be made simultaneously in two dimensions: preference and frequency. Not only could the interface acquire information concerning job preference, but it could also require that an individual indicate how frequently they liked to be engaged in a job function for which they had expressed a particular preference. Figure 1 shows an assessment screen for a single work preference, while Figure 2 shows an alternative view that is also available to an individual to make their responses. The essence of the task is that an individual can provide a self-report estimate of their work preferences in a cumulative fashion, without necessarily using numbers to express their preference (as in Figure 2's exposition).

Figure 2 shows the cumulative picture of a user's work preferences and frequencies in a two-dimensional "space" bounded by the two axes of preference and frequency. Note that at any time a user can now make adjustments in either dimension to the position of any attribute by literally moving the attributes around the display area. This screen is available at the same time as the single attribute rating screen shown in Figure 1. The position of each attribute within a bounded 0-100 axis-range two-dimensional space constitutes the "scores" for each attribute, which allows for further manipulations and relations of these attribute values with other variables, as well as coordinate structure comparisons between individuals. Current empirical estimates of short-term (five-day) test-retest reliability for this form of measurement is near 0.90. The assessment task may be tried out freely at www.staffCV.com with a complete technical exposition of the interface available at www.liv.ac.uk/~pbarrett/mariner7.htm Current research with a one-dimensional profiler for personality assessment is also described and illustrated at this Web site.

In conclusion

The definition of measurement, quantity, quantitative structure, and quantification have been described above, based upon the work and publications of Michell. What is clear from this exposition is that the nature of quantity and the definition of measurement provided by Michell is axiomatic, specific, and descriptive of measurement in the natural sciences. However, what has also been made clear is that there is no necessity for investigators in a particular area to use solely quantitatively structured

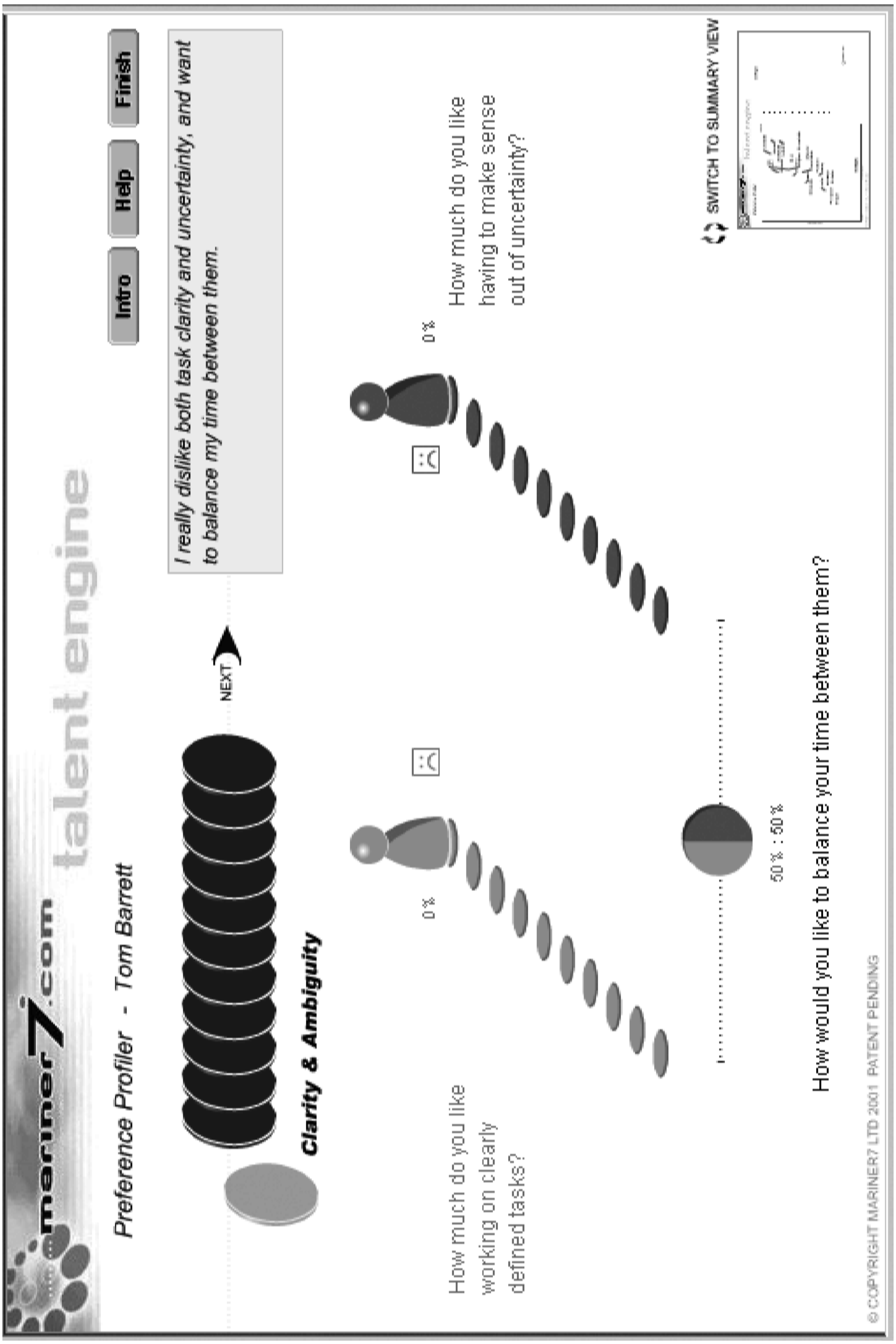

Figure 1.
The Preference Profiler single bipolar attribute assessment screen

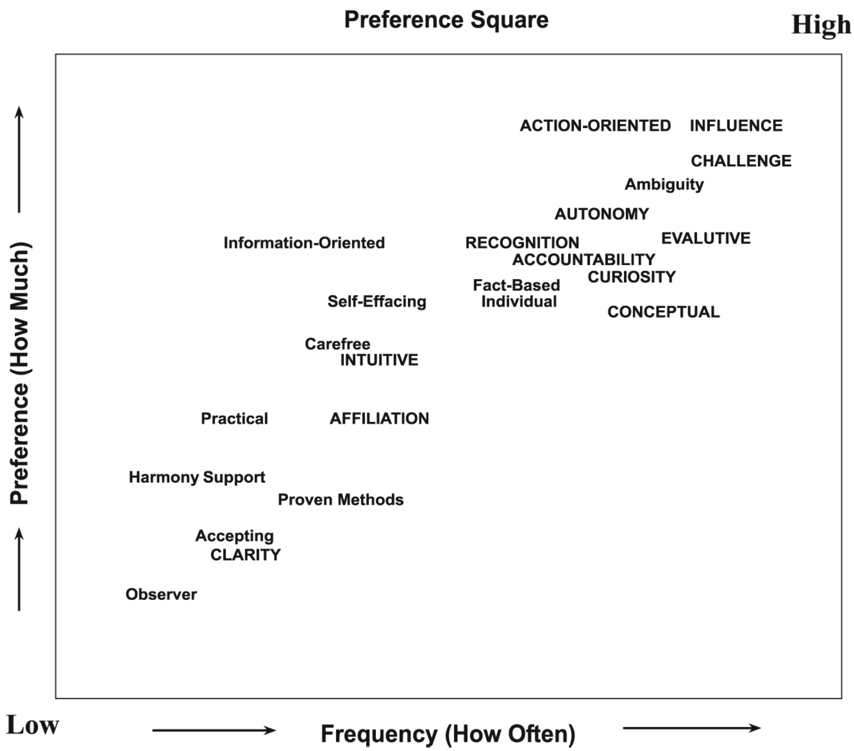

Figure 2.
The alternative format for preference assessment

variables (or operations that rely upon these) in order to justify that their investigation is scientific. That a variable might possess quantitative structure is an empirically testable hypothesis, and not necessarily the "norm" at all in psychology (as it appears to be within physics). Given much of current-day psychometrics fails to make empirical test of the quantitative structure of the variables it purports to measure quantitatively, it is concluded that it is as Michell states, a subversion of the scientific method. Looking to the future in the light of this exposition, three avenues for exploration now seem possible for psychological scientists, one that attempts quantitative measurement of psychological variables, one that attempts non-quantitative structural analysis of variables and their classifications, and one that uses the full panoply of quantitative techniques, but is careful to note that the whole exercise is approximate to some unknown degree and seeks its validity in applied predictive utility. There is no reason that activities and results from within the application of the latter two avenues cannot provide the basis for attempting to construct quantitative measurement scales for certain constructs. But, given the clear distinction between the properties possessed by a quantitatively structured variable, and those possessed by non-quantitative variables, it is hoped that a more realistic appreciation of psychological measurement and assessment may be possible by many educators, practitioners, and researchers in the area of psychological measurement. This is why the term "applied numerics" instead of "psychometrics" is suggested as a reasonable and informative description of the kinds of activities that exemplify the third and rather attractive strategy.

References

Barrett, P.T. (2001), "The role of a concatenation unit", British Psychological Society, Maths, Stats, and Computing Section Annual Conference, London, December, available at: www.liverpool.ac.uk/~pbarrett/present.htm

Barrett, P.T. (2002), "Measurement cannot occur in a theoretical vacuum", AERA Annual Educational Measurement Conference, Rasch Measurement SIG, New Orleans, LA, April, available at: www.liverpool.ac.uk/~pbarrett/present.htm

Bond, T.G. and Fox, C.M. (2001), *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*, Lawrence Erlbaum, Mahwah, NJ.

Canter, D.V. (1983), "The potential of facet theory for applied social psychology", *Quality and Quantity*, Vol. 17, pp. 35-67.

Shye, S. (Ed.) (1985), *Facet Theory: Approaches to Social Research*, Springer-Verlag, New York, NY.

Coveney, P. and Highfield, R. (1995), *Frontiers of Complexity: The Search for Order in a Chaotic World*, Ballantine Books, New York, NY.

Cronbach, L.J. (1990), *Essentials of Psychological Testing*, 5th ed., Harper Collins, New York, NY.

Donald, I. (1995), "Facet theory: defining research domains", in Breakwell, G.M., Hammond, S. and Fife-Schaw, C. (Eds), *Research Methods in Psychology*, Sage Publications, London.

Doren, D.M. (2002), *Evaluating Sex Offenders*, Sage Publications, New York, NY.

Guttman, L. (1971), "Measurement as structural theory", *Psychometrika*, Vol. 36, pp. 329-47.

Hayduk, L.A. (1996), *LISREL Issues, Debates and Strategies*, Johns Hopkins University Press, Baltimore, MD.

Hölder, O. (1901), "Die axiome der quantität und die Lehre vom Mass", *Berichte über die Verhandlungen der Königlich Sächsichen Gesellschaft der Wissenschaften zu Leipzig, Mathematisch-Physische Klasse*, Vol. 53, pp. 1-46.

Holland, J.H. (1998), *Emergence: From Chaos to Order*, Addison-Wesley, Reading, MA.

Kelley, T.L. (1929), *Scientific Method*, Ohio State University Press, Cleveland, OH.

Kline, P. (1998), *The New Psychometrics*, Routledge, London.

Kline, P. (2000), *A Psychometrics Primer*, Free Association Books, London, UK.

Krantz, D.H., Luce, R.D., Suppes, P. and Tversky, A. (1971), *Foundations of Measurement Volume 1*, Academic Press, New York, NY.

Luce, R.D. and Tukey, J.W. (1964), "Simultaneous conjoint measurement: a new type of fundamental measurement", *Journal of Mathematical Psychology*, Vol. 1, pp. 1-27.

McDonald, R.P. (1999), *Test Theory: A Unified Treatment*, Lawrence Erlbaum, Mahwah, NJ.

Michell, J. (1990), *An Introduction to the Logic of Psychological Measurement*, Lawrence Erlbaum, Hillsdale, NJ.

Michell, J. (1994), "Numbers as quantitative relations and the traditional theory of measurement", *British Journal for the Philosophy of Science*, Vol. 45, pp. 389-406.

Michell, J. (1997), "Quantitative science and the definition of measurement in psychology", *British Journal of Psychology*, Vol. 88 No. 3, pp. 355-83.

Michell, J. (1999), *Measurement in Psychology: Critical History of Methodological Concept*, Cambridge University Press, Cambridge.

Michell, J. (2000), "Normal science, pathological science, and psychometrics", *Theory and Psychology*, Vol. 10 No. 5, pp. 639-67.

Michell, J. (2001), "Teaching and misteaching measurement in psychology", *Australian Psychologist*, Vol. 36 No. 3, pp. 211-7.

Michell, J. and Ernst, C. (1996), "The axioms of quantity and the theory of measurement: part I, an English translation of Hölder (1901)", *Journal of Mathematical Psychology*, Vol. 40, pp. 235-52.

Michell, J. and Ernst, C. (1997), "The axioms of quantity and the theory of measurement: part II, an English translation of Hölder (1901)", *Journal of Mathematical Psychology*, Vol. 41, pp. 345-56.

Miles, J. (2001), *Research Methods and Statistics*, Crucial Press, Exeter.

Parsons, C. (1990), "The structuralist view of mathematical objects", *Synthese*, Vol. 84, pp. 303-46.

Perline, R., Wright, D.B. and Wainer, H. (1979), "The Rasch model as additive conjoint measurement", *Applied Psychological Measurement*, Vol. 3 No. 2, pp. 237-55.

Quinsey, V.L., Harris, G.T., Rice, M. and Cormier, C. (1998), *Violent Offenders: Appraising and Managing Risk*, American Psychological Association, Washington, DC.

Resnick, M.D. (1997), *Mathematics as a Science of Patterns*, Clarendon Press, Oxford.

Shye, S. (Ed.) (1973), *Collected Works of John Stuart Mill*, Vol. 7, University of Toronto Press, Toronto.

Shye, S. (Ed.) (1978), *Theory Construction and Data Analysis in the Behavioral Sciences*, Jossey-Bass, San Francisco, CA.

Shye, S. (1988), *Multiple Scaling*, North Holland, Amsterdam.

Stankov, L. and Cregan, A. (1993), "Quantitative and qualitative properties of an intelligence test: series completion", *Learning and Individual Differences*, Vol. 5 No. 2, pp. 137-69.

Stevens, S.S. (1951), "Mathematics, measurement, and psychophysics", in Stevens, S.S. (Ed.), *Handbook of Experimental Psychology*, Wiley, New York, NY.

Stevens, S.S. (1959), "Measurement, psychophysics, and utility", in Churchman, C.W. and Ratoosh, P. (Eds), *Measurement, Definitions, and Theories*, Wiley, New York, NY.

Suen, H.K. (1990), *Principles of Test Theories*, Lawrence Erlbaum, Hillsdale, NJ.

Swets, J.A., Dawes, R.M. and Monahan, J. (2000), "Psychological science can improve diagnostic decisions", *Psychological Science in the Public Interest*, Vol. 1 No. 1, pp. 1-26.

Thomson, W. (1891), *Popular Lectures and Addresses*, Vol. 1, Macmilllan, London.

Thorndike, E.L. (1904), *Theory of Mental and Social Measurements*, Science Press, New York, NY.

Thorndike, E.L. (1918), "The nature, purposes, and general methods of measurements of educational products", in Whipple, G.M. (Ed.), *Seventeenth Yearbook of the National Society for the Study of Education*, Vol. 2, Public School Publishing, Bloomington, IL, pp. 16-24.

Ulam, S. (1952), "Random processes and transformations", *Proceedings of the International Congress on Mathematics*, Vol. 2, pp. 264-75.

Vion, D., Aassime, A., Cottet, A., Joyez, P., Pothier, H., Urbina, C., Esteve, D. and Devoret, M.H. (2002), "Manipulating the quantum state of an electrical circuit", *Science*, Vol. 296 No. 3 May, pp. 886-9.

Wilson, M. (1995), "Structuring qualitative data: multidimensional scalogram analysis", in Breakwell, G.M., Hammond, S. and Fife-Schaw, C. (Eds), *Research Methods in Psychology*, Sage Publications, London.

Wood, R. (1978), "Fitting the Rasch model: a heady tale", *British Journal of Mathematical and Statistical Psychology*, Vol. 31, pp. 27-32.

Wolfram, S. (1994), *Cellular Automata and Complexity: Collected Papers*, Addison-Wesley, Reading, MA.

Wolfram, S. (2002), *A New Kind of Science*, Wolfram Media, Inc., New York, NY.