Beyond Psychometrics: Measurement, non-quantitative structure, and applied numerics

Author: Paul Barrett
(Augmented Web Version)

Affiliations

Chief Psychologist
Mariner7 Ltd.
640 Great South Road
Private Bag 92-106
Manukau
Auckland
New Zealand

Hon. Associate Professor
University of Auckland
Dept. of Psychology
Symonds Street
Private Bag 92019
Auckland
New Zealand

Hon. Associate Professor
University of Canterbury
Dept. of Psychology
Christchurch
Private Bag 4800
New Zealand

Hon. Senior Research Fellow
University of Liverpool
Dept. of Clinical Psychology
Whelan Building
Brownlow Hill
Liverpool
United Kingdom

Contacts

Tel: +64-9-262-6082
Fax: +64-0-262-6290

Email: paul.barrett@mariner7.com and paul.barrett@auckland.ac.nz

*Shortened paper published as:

Barrett, P.T. (2003) Beyond Psychometrics: Measurement, non-quantitative structure, and applied numerics. *Journal of Managerial Psychology*, 3, 18, 421-439.

Abstract

A simple statement from Michell (2000) … "psychometrics is a pathology of science" is contrasted with the content of conventional definitions provided by leading textbooks in the area. The key to understanding why Michell has made such a statement is bound up in the precise 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 in both classical and item-response-theory, it is clear that Michell is indeed correct in his assertion. Three avenues of investigation would seem to follow from this understanding: each of which is expected to gradually replace current psychometric test theory, principles, and properties. The first attempts to construct variables which can be demonstrated empirically to possess a quantitative structure, and then use these for applied and theory-based measurement. 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 put to one side in favour of simpler tests of observational reliability and validity. Examples are presented of what "practice" now looks like in each of these avenues. Where many of the 20th century developments in psychometrics were mainly concerned with finding novel ways to manipulate and work with numbers and test scores, it is expected that psychologists in the 21st century will begin to recognise that the "quantitative imperative" (Michell (1990) is not necessary to the scientific study of psychology. Further, where variables are sought to be quantified, it will be recognized that this "quantification" requires an explicit hypothesis to be tested, prior to the subsequent manipulation of any variable magnitudes by operations that rely upon an additively structured variable. It is to be hoped that psychology begins concerning itself more with the logic of its measurement than the ever-increasing complexity of its numerical and statistical operations.

Consider the following statements from Michell (2000) p. 639 … "It is concluded that psychometrics is a pathology of science…" and Michell (2001), p. 211 … "the way in which psychometrics is currently, typically taught actually subverts the scientific method". Now consider the following definitions of psychometrics from a sample of current textbooks: Kline (2000), page 1, defines psychometrics as … "Psychometrics 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 … "Psychometric 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), page 4, defines it as … "The 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 … "Test theory is an abbreviated expression for *theory of psychological tests and measurements*, which in turn can be abbreviated back to *psychometric theory* (psychological measurement)". Finally, Miles (2001), p. 62, defines psychometrics as … "Psychometrics is 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 charges that psychometrics is a pathology of science. It is easy to dismiss Michell's writings as just another occasional outburst by a disaffected academic or the usual periodic surfacing of criticism of the status quo in an established field of psychological enquiry. However, Michell's logic is inexorable, leading to just those conclusions he has espoused. Let me briefly adumbrate that logic, and the bases upon which it is constructed, using, where desirable, relevant passages from various of Michell's publications.

Psychometrics is indeed concerned with the measurement of psychological attributes. These attributes are non-observable, inferred, hypothesised variables who existence is to be inferred through the measurement and manipulation (where possible) of other variables and their theoretically expected relations amongst one another. But, the focal point is in the meaning of that word *measurement*. It is not the "catch-all" term that most psychologists seem to think. There are four critical points of understanding to be addressed:

1: Quantitative Measurement

Michell (1990), p.63 ...

" Quite simply, measurement is a procedure for identifying values of quantitative variables through their numerical relationships to other values. Take a simple example. We wish to know the length of a timber beam. This may be done by relating its length to that called a meter. It is to be found r meters long (where r is some real number). Here r is the ratio of the length of the beam to that of a meter and this FACT enables the length of the beam to be characterized. More generally, in measurement some (unknown) value of a quantitative variable is identified as being r units. A UNIT of MEASUREMENT is simply a particular value of the relevant variable. It is singled out as that value relative to which all others are to be compared. Let the unit be Y and let the value to be measured be X. Then a measurement has the form X = rY.... Measurement requires the development of procedures whereby values X and Y may be brought into comparison and their ratio assessed. Such procedures are the methods of measurement"

Michell (2001), p. 212 …

"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 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".

What is immediately apparent is that 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**". Yes, this is a narrow definition for measurement, but it is unambiguous and technically specified as we shall see below.

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. Specifically (from Michell, 1990), p. 52-53) …

> "The first fact to note about a quantitative variable is that its values are ordered. For example, lengths are ordered according to their magnitude, 6 meters is greater than 2 meters, and so on. Similarly the values of other quantitative variables are ordered according to their magnitudes. The familiar symbols, "≥" and ">" will be used to denote this relation of magnitude, "≥" meaning "at least as great as", and ">" meaning "greater than". Also the symbol "=" will be used to signify identity of value.
>
> Let X, Y, and Z be any three values of a variable, Q. Then Q is **ordinal** if and only if:
>
> 1) if $X \geq Y$ and $Y \geq Z$ then $X \geq Z$ (transitivity)
>
> 2) if $X \geq Y$ 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 ≥ is a simple order on its values. All quantitative variables are simply ordered by ≥ , 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 ≥ Y if and only if X + Y ≥ 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 *n*X ≥ Y (where 1X = X and (*n* + 1)X = *n*X + X) (the Archimedean 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, and later by Hölder (1901) 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, this is not some piece of ad-hoc philosophy produced to support a convenient argument, but rather, these are the bases for the kind of quantitative measurement that has evolved within the natural sciences.

3. Numbers and their status

Up to now, it has been possible to regard the properties of measurement in isolation of the numbers used to represent magnitudes. However, this third issue is also fundamental to an understanding of measurement. It is also perhaps the key to understanding measurement in its wider context. 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 as in section 2 above. 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 2. above 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 1000, 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. In fact, the same relations as expressed in 2. above can also be applied to such a number system (all positive integers). 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.

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 in the manner of Stevens (1951) theory, p. 23 …

> "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. The conventional series of numerals – the series in which by definition each member has a successor – yields to analogous operations: We can identify members of the series and classify them. We know their order as given by convention. We can determine equal differences, as 7-5=4-2 and equal ratios, as 10/5 = 6/3. 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 (p. 19) … "provided a consistent rule is followed, some form of measurement is achieved".

This seems a reasonable statement on the surface, and it is has taken the form of a mantra chanted by all undergraduate psychology students worldwide. But, it is deeply flawed. What Stevens did was to remove the status of a numerical relation system consisting of the real numbers as an empirical system in its own right. Up until the 1950s, numbers were considered to constitute an empirical relational system in their own right. The system was self-contained, logical, possessed the required ordering relations that constitute both ordinal and additive operations, and, in the theory of continuous quantity, sustained the necessary ratios necessary for such a theory. In short, both in the manner that scientists used them, as well as in their existence as a relational system, numbers were considered as empirical facts, not abstract entities. The existence of the empirical relations was presumed logically independent of the numerical assignments made to represent them. 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**. When stated like this, it is obvious to even the most disbelieving reader that this is not how measurement in the natural sciences has ever functioned – neither is it a rational course of action for constructing and making measurement.

When 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…
>
> "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. A science that aspires to be quantitative will ignore this fact at its peril. It is pointless to invest energies and resources in the enterprise of quantification if the attribute involved is not really quantitative. The logically prior task in this enterprise is that of addressing this empirical issue. I call it the scientific task of quantification (Michell, 1997)".

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 6th 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 17th century. As Michell (2000) p. 653 puts it:

> "The scientists of the 17th century measured what they could, attempted to make measurable what they could not, and what they could not measure, they doubted the reality of. Attributes found to be measurable they thought of as *primary qualities*. The

> remainder they called *secondary qualities*... The operational distinction, based in measurement, between primary and secondary qualities was transformed by Descartes into a metaphysical distinction between separate realms of being, those of body and mind. Mental phenomena were excluded from science because they were excluded from quantity."

With the success of quantitative physics in the 19th century, came an almost absolute certainty that what could not be measured was of no substantive scientific import. The Kelvin dictum was born during this century (Thomson, 1891, p.80-81) ...

> "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 the Thorndike Credo of 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 in 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 17th through 19th 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. Since psychology, both for academic and financial credibility, needed to advertise itself as a science; it subsequently adopted the procedures and practices of quantitative measurement as found within the natural sciences. However, the *quantitative imperative* (Michell, 1990) was based upon two false premises: firstly that in order for any area of investigation to be considered a science, it must use quantitative measurement of its variables, and second, that all variables in psychology were quantitatively structured. 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"

Given the four critical points above, it is clear that Michell's use of the word "measurement" is concordant with the axioms of quantity, in that variables so measured possess both ordinal and additive ordered structures, with the appropriate ordinal and additive structured numerical system used to "represent" the empirically defined object properties. That there is little disagreement with the above is testament to the veracity of both the axioms and the status of numbers as empirical facts, within a logically independent empirical relational system. However, if we apply this logic to the kind of variables used routinely in psychology, such as personality traits, intellectual abilities, IQ, preference judgements, attitudes etc., it is clear that as yet, little empirical evidence exists for any of them being structured as quantitative variables. What little there is has been explicitly tested using the conjoint measurement axioms of Luce and Tukey (1964), which will be discussed below.

When confronted with this fact, for it is a fact, many psychologists retort that psychological measurement "is different from" measurement in the natural sciences. When pressed to explain the new axiomatic basis (or specific conditions) for this special measurement in psychology, there is complete silence. The issue here for many in psychology is not so much that Michell may be wrong in his exposition of the theory of measurement and continuous quantity, but whether what he states is in any way relevant to psychological and psychometric measurement. However, this "relevance" question is itself based upon a false premise. That is, that there exist different kinds of quantitative measurement which are relevant to particular domains of enquiry. There are not. The axioms defining quantity and the theory of continuous quantity that underlines quantitative relations and structures is not an "option", but possess the status of empirical facts. What is questionable though is whether explanatory variables proposed in psychology possess a quantitative structure such that they can be quantified in the manner of a natural science. This is the empirically testable "scientific hypothesis" to which Michell (1997) refers.

The strongest statement rejecting Michell's thesis was published by Lovie (1997), in response to Michell's (1997) paper. Lovie states (p. 393) ..

> "there are no absolute, ahistorical mathematical truths or methods, only locally developed and locally maintained collective commitments and practices; what the ethnomethodologist Eric Livingston has termed the 'lived work' of the practising mathematician (Livingston, 1986)."

As Michell (1999) details, the definition of measurement and the process of quantification outlined in the 4 critical points above stems from Euclid onwards. Both Newtonian and the New Physics, let alone chemistry and biology are predicated upon these quantity axioms. This work and knowledge constitutes a human-race-wide effort. If this is what Lovie meant as "locally maintained" and "collective", it is clear that his criticism is actually no criticism at all. However, it is apparent from the remainder of his critique that Lovie really does mean that the axioms of quantity are constructivist "entities" – of no particular relevance to one area of investigation than to another, except that within which they are "maintained" by the investigative "collective". So, how psychologists as a "collective" wish to define measurement and quantity is entirely up to them. The problem for Lovie is that whilst refusing to accept the axioms of quantity, like so many other psychologists who do the same, he is quite unable to provide any other definition of quantity. Instead, it seems to be that whatever is said to constitute a "collective" is responsible for whatever definition (or not) they wish to propose. This will not do. The axioms above represent mankind's historical formalisation of what it has been engaged in for thousands of years. We all implicitly use these properties of numbers in our everyday lives. Our technologies and our very lives are constructed around these properties of measurement. But, psychologists seem able to decide that this "kind of" measurement is not for them, instead preferring "something else" without ever making explicit that which they practise. Well, this paper makes it explicit for them. It is *applied numerics*, not quantitative measurement. As a group they are entirely free to use whatever definition of measurement they wish, or even not to have one at all, but they cannot at the same time claim to be making quantitative measurement of psychological attributes, or make claims about how variables interact with one another or cause certain outcomes simply by using the numeric techniques of quantitative science.

Note that 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 invariably touted by psychologists as an exemplar for “a different kind of measurement” or at least a “look how physics has changed” kind of statement), 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, Aassime, Cottet, Joyez, Pothier, Urbina, Esteve, and Devoret, 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 whilst the quantitative structure of the variables so manipulated remains untested. As Michell (1986) and later in 1999, p. 45 & 46) stated ...

> "It can be seen that the calculation of means of ordinal scale measurements is generally not helpful in scientific research. There is nothing to stop one from doing it, and any conclusions arrived at will be just as empirically meaningful as any other conclusions one arrives at in scientific investigation. It is just that considering only the empirical data on which the ordinal scale is based, no empirical conclusions about means validly follows. To compute the mean is to go beyond the data given, and to infer empirical conclusions about it, is to infer what cannot validly follow from that data."

Those who use such quantitative methods, drawing conclusions based upon the real continuous number manipulations and unit-preserving operations that are involved in such techniques, are committing a logical error of such magnitude that it is little surprise that so little of this work is replicable, let alone scientifically valuable.

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… Thorndike (1904), p. 19 …

> "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"

However, such "measurement" is 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? Remember that 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. In order to clarify the importance of this final point, look first at the quote from Kelley (1923), p. 418,

> "It might seem axiomatic that there cannot be a science of quantitative measurement until and unless there is established a particular unit of measurement. This is, however, true only in a limited sense; for it is quite conceivable that one could have a science of physical phenomena to which the units were such that the scale of time intervals was the square of the present intervals measured in seconds, and in which the length scale was logarithmic as compared to the present scale in centimetres etc. Of course, in terms of these new units, all the laws of physics would be stated by means of formulas different from and in general more cumbersome than our present formulas; but, nevertheless we could have an exact science. The existence of science does not lie in the units employed, but in the relationships which are established as following after the choice of units".

and then Michell's (1999), p. 105 response …

"Thorndike's problem had been that because items in a mental test might differ in level of difficulty, the observed score is a sum of units of different magnitude. This had led him to prefer measurement by relative position, which Boring was now classing as mere 'rank order'. Kelley's retort was to draw attention to a very subtle and not widely appreciated degree of freedom within quantitative science. His observation about transforming scales for measuring time and length is quite true. What he failed to bring out, however, is that it is really quite a different problem from that facing psychologists. The previous chapter drew attention to the fact that for any two magnitudes of a quantity (say any two lengths) there is no unique ratio between them. Ratios are tied to relations of additivity. If for any continuous attribute there is one such relation, then there is an infinite number. Replacing our conventional scale of length (which is based upon our conventional view of what it is for lengths to add together) by one which is its logarithmic transform, as Kelley suggests, simply identifies a different relation of additivity between lengths, one which although it seems quite unnatural to us, exists alongside the other. Physics has the luxury of being able to select whichever additive relations best suit (as the case of velocity illustrates), but this is a luxury bestowed in virtue of already having discovered that its attributes possess additive structure. There is no parallel here with the situation then existing in psychometrics and there could be none until it is shown that attributes like ability or knowledgability are quantitative."

It is now hoped that the reader can begin to see the sheer illogic in much of what constitutes psychometric and psychological measurement, and why many in psychology continue to persevere in maintaining that psychometric measurement **must be** concordant with quantitative measurement. That there is no such imperative is now clear. 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 facts 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.

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 avenue 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). Thirdly, 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, 2 and 3-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. In 1964, Luce and Tukey 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. Whilst 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, Luce, Suppes, and Tversky (1971) have since provided the complete set of formal proofs for the conjoint measurement axioms. Hölder (1901) had initially provided the logic of indirect tests of quantitative structure, utilising theorems concerning the additive composition of intervals on a straight line. For example, given two intervals on a line, from point A to point C (AC), and point D to point F (DF), and given two intermediate points B (within the AC interval) and E (within the DF interval), then given AB= DE and BC=EF, then the distance between AC must equal the distance between DF if the units of length are additive on the straight line. The proposition is that what must be true for intervals on a straight line must also be true of differences within any such quantitative attribute. Luce and Tukey generalised this logic to combinations of attributes in a scenario which enabled differences within two attributes to be matched between them relative to their joint effects on a third attribute. 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.

Given LS_{score} = Letter series test score

Intelligence ≡ Letter Series correct completion

M = the Motivation variable

WM = the Working Memory variable

M_{score} = Motivation condition (ordinally increasing levels of motivation required)

WM_{score} = Working memory score (working memory place-keepers)

Then, assuming the variable LS_{score} possesses a theoretically infinite number of values, the three key initial conditions for conjoint measurement are:

Intelligence = f(M, WM); that is, Intelligence is some mathematical function of Motivation and Working Memory.

There is a simple order (the relation ≥) upon the values of Intelligence (LS_{scores} are ordinally related)

The values of M amd WM (**M_{score} and WM_{score}) can be identified (i.e. objects may be classified according the the values of M and WM they possess).** Further, they can be manipulated independently of each other. That is, the values for M and WM can be realized independently from one another.

The logic of the procedure for assessing whether Intelligence is a quantitatively structured variable is as follows:

Assume persons P_1 and P_2 obtain the same LS_{score}, but they differ in the amounts of M and WM (as indexed by the M_{score} and WM_{score}s). P_1 has a higher M_{score} than P_2, but P_2 has a higher WM_{score} than P_1. What is being tested is the functional relation: Intelligence = M + WM. If this additive relation holds, then the differences between M_{scores} for P_1 and P_2 = the differences between WM_{scores}. The basic idea is that levels within either of the two attributes (M and WM) can be traded off against one another relative to the effects on the Intelligence variable. By acquiring values of Intelligence, W and WM (as LS_{score}, M_{score}, and WM_{score}) and comparing these values in the manner required to test the conditions for conjoint additivity, it is possible to **empirically determine** whether an unobserved, latent variable (such as intelligence) is indeed quantitatively structured.

Of critical importance is the realisation that Rasch item response theory is also an empirical instantiation of the conjoint additivity axioms (Perline, Wright, and Wainer, 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. 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. Wright (1999) also provides a clear, succinct, and non-technical summary of the entire history and rationale behind the evolution of the Rasch model. For a more philosophical treatise on the essence of objectivity in measurement, Fisher's (1992) chapter is essential reading. The Institute for Objective measurement in Chicago, is already a reality and has been so for many years. This Institute is devoted to the theory, procedures, and methods for the construction of quantitative measures. Its many members are routinely producing such measures within a wide domain of investigation, from medicine, education, sociology, through to psychology. Within the state of Florida, the Rasch constructed Lexile unit scale has been used as a standard measure of reading proficiency for many years now. From Stenner, the creator of the Lexile scale (personal communication)...

> "In the Lexile Framework for Reading (www.lexile.com) item calibrations come from theory and these calibrations embody our intentions regarding the reading variable independent of the person response data. Person fit is at once a test of the quantitative hypothesis (Michell, 1999) and the substantive construct theory. Good fit over 10,000's persons, different item formats and different demographic and age groupings means that the Lexile Theory tells a useful story about what reading is".

Whilst the construction of variables that possess quantitative structure is now possible within psychology, *a-priori* meaning instantiation remains critical. As Barrett (2001 and 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 psychologists 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. What should be slowly becoming clear from the above statements is that quantity is not synonymous with mathematics. If mathematics is considered as the science of abstract structure (as indicated earlier), 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, with the article title "Measurement as structural theory". 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 extremely simple introductions to this area of research, whilst Canter (1983, 1985) provides a thoroughgoing exposition of facet theory. Much of this work now takes place within the domain of offender profiling research, with Canter the UK's leading exponent of facet theory and what is called "investigative psychology". To give the reader the flavour of Guttman's approach to psychology, his statement in 1991, p. 42 makes his position clear…

> "Those who firmly believe that rigorous science must consist largely of mathematics and statistics have something to unlearn. Such a belief implies the emasculation of the basic substantive nature of science. Mathematics is content-less, and hence not -in itself- empirical science … rigorous treatment of content or subject matter is needed before some mathematics can be thought of as a possibly useful (but limited) partner for empirical science".

This view is absolutely concordant with that of Michell. Facet theory has proven to be an extremely versatile and powerful means of relating psychological theory to empirical analysis of data structures. In essence, it is a meta-theoretical approach to empirical research, based in set theory terms, and deals with membership and classes rather than point-estimates on linear additive scales of measurement. Fifty years of research has demonstrated both its utility and credibility. The fact that it has not been used more as a means of investigation is again due to the quantitative imperative that many psychologists find impossible to avoid, alongside the practicalism that demands that almost every observation be reduced to a number or statistic for pragmatic convenience.

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 which 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 in 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. These "lattice" models are now used routinely for fluid dynamics, porosity dynamics and cement hydration). However, such systems (the study of the evolution of artificial life being one such domain of investigation) do seem to mimic certain real-world phenomena to high degree of congruence. This kind of work is maintained as a coherent research strategy at the Santa Fe Institute in the US (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 actually possess these properties of quantity is never tested. It is within this avenue in which classical and modern 2 and 3-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 within here. Whilst 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 2 and 3-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 no attention is ever 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. However, it is worth examining in detail the justification for this from at least one exponent of structural equation modelling. In a public debate with this author on measurement issues via SEMNET, a professional email listserv group that discusses issues concerned with structural equation modelling and whose message archives can be searched at http://bama.ua.edu/archives/semnet.html) Hayduk (29th May, 2002) has responded to a quote from Michell (1990), p.63 last paragraph ...

> "Having clarified these preliminary issues the meaning of measurement becomes obvious. Quite simply, measurement is a procedure for identifying values of quantitative variables through their numerical relationships to other values"

with Hayduk's response as:

> "I find major fault with Michell's definition in that it is ambiguous with respect to the necessary presence of the "world out there" as the "stuff" being measured. Some prior, or presumed, or assumed, feature of the world is being measured. Michell might have been intending to squash the whole world into his word "variables" but I think not. Just try reading this as "procedures for identifying values of quantitative variables existing in the world yet known to us only imperfectly and unclearly since we do not yet possess any clean/clear/infallible understanding of that world..." This would raise issues Michell does not seem to want to address, and probably can not address, yet which must be addressed if one is to speak of measuring features of the world out there. A supposed definition of measurement that fails to centrally incorporate the notion of the "stuff" "features"

"structure" "shades" "noticeable-progressions" of the world out there, is not a definition SEM can abide/condone. SEM latents are stand-ins for, or representations of, or characterizations of, that world out there. In SEM measurement is the structured connection BETWEEN that world and the indicators, and measurement is NOT merely a property or properties of the indicators themselves. SEM's notion of measurement demands a central place for the featured world, and Michell's definition fails to incorporate the featured world as essential"

In response to page 75 of Michell (1999) ...

"Because measurement involves a commitment to the existence of quantitative attributes, quantification entails an empirical issue: is the attribute involved quantitative or not? If it is, then quantification can sensibly proceed. If it is not, then attempts at quantification are misguided. A science that aspires to be quantitative will ignore this fact at its peril. It is pointless to invest energies and resources in the enterprise of quantification if the attribute involved is not really quantitative. The logically prior task in this enterprise is that of addressing this empirical issue. I call it the scientific task of quantification."

Hayduk replies ...

"No this task is NOT logically prior. The appearance of the latent within the latent level model is what tells us as SEM researchers that there may well be a latent that EXISTS due to its reasonable/understandable connection to a web of other latents in the model. This evidence of the existence of the latent comes along with, accompanies, is necessarily-part of, the discussion of the connection between the latent and the indicators. The claim to logical prior-ness here is merely Michell's blindness with respect to the need for a worldly entity being required. If Michell kept the world in mind, he would not be able to claim logical prior-ness here. Measurement is inextricably bound to, and mixed with, hidden among, our conceptualizations of multiple things/entities/latents and all the procedural stuff that is done as the methods of data collection. Measurement can not be separated out as if it stands apart from our latent-level conceptualization (even if biased conceptualization) of the world out there".

What is apparent from the above two responses from Hayduk is that he sees measurement within structural equation modelling as "something different" from that as defined by Michell. However, there is a fundamental misunderstanding that is prevalent throughout these passages, common to many psychologists who reject Michell's statements. This is that Michell's thesis and the axiomatic basis of quantitative measurement is viewed as somehow disconnected from some notion of "real world stuff", such that the definition for quantity and theory of continuous quantity is marginalised in order that the investigator can proceed with the task of "making sense of the world out there". However, mischaracterising Michell is no answer to the issues above. Note the basis for measurement is the conjoining of an empirical entity relational system with that of a numerical relational system. The empirical relational system (whether including latent variables or otherwise) is required to be investigated or defined independently of the use of any number system. Where a variable is unobservable (non-physical), then the empirical task becomes one of assessing whether a theoretically proposed mapping of numbers (which possess additive relations) onto the hypothetical quantities of the latent variable is justified. Additive conjoint measurement theory achieves just that task. Hayduk instead proposes that a model network of variables and additive relations, imposed as an *a-priori* set of measurement and relational statements, is also sufficient to assure an investigator that the variables used within such a model must necessarily possess quantitative structure, **if** the model fits an expected "population" covariance matrix generated from the observed data covariances. At first glance, this approach seems reasonable, for surely, if a model fits the maximum likelihood estimated population covariance data, then this must indicate that measurement has been achieved in the manner defined (all variables possess both ordinal and additive relations between their values)? The problem with this approach is that it confuses measurement with model fit. It is possible to model relations between quantitative variables, yet still achieve no-fit, because the model inappropriately specifies how these variables are causal for some outcome/s. Likewise, it is possible to model with ordinal-relation variables that are assigned numerals for each of their amounts, treat the numerals as though they represented the actual quantitative amounts of the latent variables involved, then obtain a model-fit to the population covariance data. For example, we might achieve fit with variables such as extraversion, self-esteem, religiosity etc., and so conclude that these variables now possess quantitative structure, yet, the quantitative structure actually resides within the numerical relational system and not necessarily the empirical relational system. The empirical relational system has never in fact been examined. Of course, it is always possible that the investigator has guessed right – and that model fit does indeed indicate that all variables possess a quantitative structure. The point being that fitting SEM models cannot test the empirical hypothesis of quantitative variable structure as SEM's arithmetic operations are constructed on the prior assumption that all variables **must be** quantitative from the outset. In fact Hayduk's position looks remarkably similar to the credo from Cronbach and Meehl (1955) concerning construct validity…

> "Scientifically speaking, to 'make clear what something is' means to set forth the laws in which it occurs."

This is akin to Hayduk's justification of modelling real world stuff with SEM, and that model-fit implies that one better understands the phenomena being modelled. However, note Maraun's (1998), p. 448, response to the Cronbach and Meehl statement …

> "This is mistaken. One may know more or less about *it*, build a correct or incorrect case about *it*, articulate to a greater or lesser extent the laws into which *it* enters, discover much, or very little about *it*. However, these activities all presuppose rules for the application of the concept that denotes *it* (e.g. intelligence, dominance). Furthermore, one must be prepared to cite these standards as justification for the claim that these empirical facts are about *it*…the problem is that in construct validation theory, **knowing** about something is confused with an understanding of the **meaning** of the concept that denotes that something".

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 mistake 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.

Whilst 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 well-being, 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, Harris, Rice, and Cormier, 1998; Doren, 2002) and in the monograph by Swets, Dawes, and Monahan (2000) on making diagnostic decisions using signal detection theory.

Within an organizational 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 already a reality. This is the preference profile™ technology currently marketed by Mariner7 Ltd. 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 Mariner7's 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, whilst Figure 2 shows an alternative view which 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 1 and Figure 2 here

Figure 2 shows the cumulative picture of a user's work preferences and frequencies in a 2-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 2-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 (5-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 website.

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 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 from: 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, April. Available from: 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*. Mahwah, New Jersey: Lawrence Erlbaum.

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

Canter, D. V.(ed) (1985) *Facet Theory: Approaches to Social research*. New York: Springer-Verlag.

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

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

Cronbach, L.J., and Meehl, P. (1955) Construct validity in Psychological Tests. *Psychological Bulletin*, 52, , 281-302.

Donald, I. (1995) Facet Theory: Defining Research Domains. In G.M. Breakwell, S. Hammond, and C. Fife-Schaw (eds.) *Research Methods in Psychology*. London: Sage Publications.

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

Fisher Jnr., W.P. (1992) Objectivity in Measurement: A Philosophical History of Rasch's Separability Theorem. In Wilson, M. (ed). *Objective Measurement: Theory into Practice*. Norwood, New Jersey: Ablex Publishing.

Guttman, L. (1971) Measurement as structural theory. Psychometrika, 36, 329-347.

Guttman, L. (1991) *Chapters from an Unfinished Textbook on Facet Theory*. Jerusalem: Hebrew University Press.

Hayduk, L.A. (1996) *LISREL issues, debates and strategies*. Baltimore: Johns Hopkins University Press.

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*, 53, 1-46 (translated in Michell and Ernst, 1996, 1997).

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

Kelley, T.L. (1923) The principles and technique of mental measurement. *American Journal of Psychology*, 34, 408-432.

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

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

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

Livingston, E. (1986) *The Ethnomethodological Foundations of Mathematics*. London: Routledge and Kegan Paul.

Lovie, A.D. (1997) Commentary on Michell, Quantitative Science and the definition of measurement in psychology. *British Journal of Psychology*, 88, 393-394.

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

Luce, R.D., Krantz, D.H., Suppes, P., and Tversky, A. (1989) *Foundations of Measurement Vol. 3: Representation, Axiomatization, and Invariance*. New York: Academic Press.

Maraun, M.D. (1998) Measurement as a Normative Practice: Implications of Wittgenstein's Philosophy for Measurement in Psychology. *Theory & Psychology*, 8(4), , 435-461.
\
McDonald, R.P. (1999) *Test Theory: a Unified Treatment*. Mahwah, New Jersey: Lawrence Erlbaum.

Michell, J. (1986) Measurement scales and statistics: a clash of paradigms. *Psychological Bulletin*, 100, 3, 398-407.

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

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

Michell, J. (1997) Quantitative science and the definition of measurement in Psychology. *British Journal of Psychology*, 88, 3, 355-383.

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

Michell. J. (2000) Normal science, pathological Science, and psychometrics. *Theory and Psychology*, 10, 5, 639-667.

Michell, J. (2001) Teaching and misteaching measurement in psychology. *Australian Psychologist*, 36, 3, 211-217.

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*, 40, 235-252.

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*, 41, 345-356.

Mill, J.S. (1983) *A system of logic*. London: Parker.

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

Parsons, C. (1990) The structuralist view of mathematical objects. *Synthese*, 84, 303-346.

Perline, R., Wright, D.B., and Wainer, H. (1979) The Rasch Model as Additive Conjoint Measurement. *Applied Psychological Measurement*, 3:2, , 237-255.

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

Resnick, M.D. (1997). *Mathematics as a science of patterns*. Oxford: Clarendon Press.

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

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

Stankov, L. and Cregan, A. (1993) Quantitative and qualitative properties of an intelligence test: series completion. *Learning and Individual Differences*, 5, 2, 137-169.

Stevens, S.S. (1951) Mathematics, measurement, and psychophysics. In S.S. Stevens (ed). *Handbook of Experimental Psychology*. New York: Wiley.

Stevens, S.S. (1959) *Measurement, psychophysics, and utility*. In C.W. Churchman and P. Ratoosh (eds.) Measurement, Definitions, and Theories. New York: Wiley.

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

Swets, J.A., Dawes, R.M., Monahan, J. (2000) Psychological Science Can Improve Diagnostic Decisions. *Psychological Science in the Public Interest*, 1, 1, 1-26.

Thomson, W. (1891) Popular Lectures and Addresses Vol. 1. London: MacMilllan.

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

Thorndike, E.L. (1918). The nature, purposes, and general methods of measurements of educational products. In G.M. Whipple (ed.), *Seventeenth yearbook of the National Society for the Study of Education, Vol. 2* (pp. 16–24). Bloomington, IL: Public School Publishing.

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*, 296, 3rd May, 886-889.

Wilson, M. (1995) Structuring Qualitative Data: Multidimensional Scalogram Analysis. In G.M. Breakwell, S. Hammond, and C. Fife-Schaw (eds.) *Research Methods in Psychology*. London: Sage Publications.

Wolfram, S. (1994) *Cellular Automata and Complexity: collected papers*. Reading, Massachusetts: Addison-Wesley.

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

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

Wright, B.D. (1999) Fundamental Measurement for Psychology. In S.E. Embretson and S.L. Hershberger (eds.) *The New Rules of Measurement: What Every Psychologist and Educator Should Know*. Mahwah, New Jersey: Lawrence Erlbaum.

Figure 1: The Preference Profiler single bipolar attribute Assessment Screen

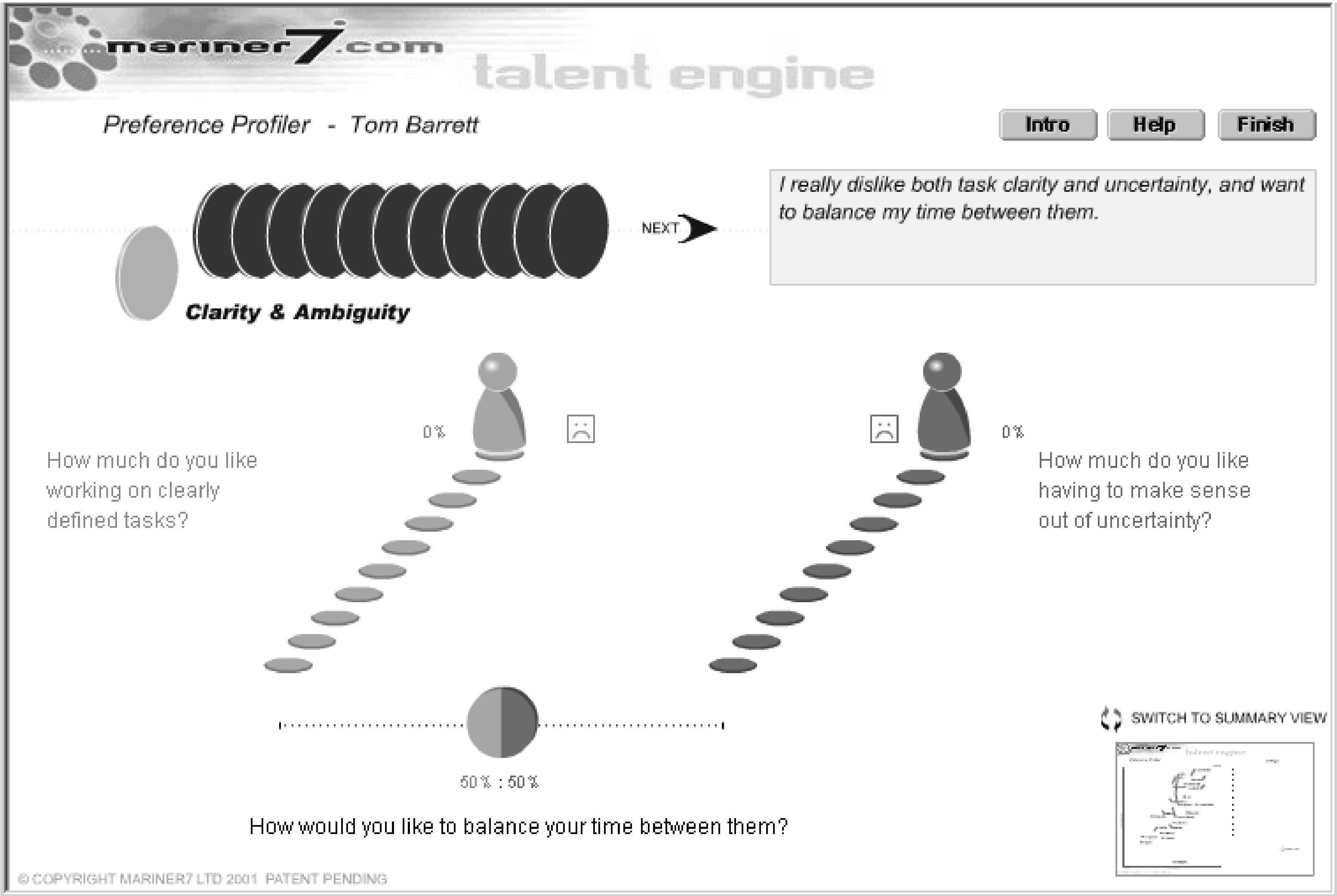

Figure 2: The alternative format for preference assessment