Chapter 1B

Reducible Or Irreducible? Mathematical Reasoning and the Ontological Method

William P. Fisher, Jr.
MetaMetrics, Inc.

Science is often described as nothing but the practice of measurement. This perspective follows from longstanding respect for the roles mathematics and quantification have played as the media through which alternative hypotheses are evaluated and experience becomes better managed. Many figures in the history of science and psychology have contributed to what has been called the "quantitative imperative," the demand that fields of study employ number and mathematics even when they do not constitute the language in which investigators think together. But what makes an area of study scientific is, of course, not the mere use of number, but the existence of a community of investigators who share a common mathematical language for exchanging quantitative value. Such languages require a rigorous theory for each variable involved, and a reference standard quantitative metric to which each instrument brand or configuration must be equated. The values shared and exchanged by such communities typically involve the application of mathematical models that specify the sufficient and invariant relationships necessary for rigorous theorizing and instrument equating. This presentation explores the mathematical metaphysics of science with the aim of connecting principles of quantitative measurement with the structures of sufficient reason.

The interpretability of psychological measures

Science is often described as nothing but the practice of measurement. This perspective follows from longstanding respect for the roles mathematics and quantification have played as aids in evaluating alternative hypotheses and managing experience. Quantification was for centuries uncritically accepted as the hallmark of science, and often still is, with many historical figures contending that:

> the "world is number" (Pythagoras);
>
> in Platonic fashion, "all causes of natural effects can be discovered by lines, angles, and figures" (Grosseteste);
>
> the "book of nature ... is written in mathematical language" (Galileo);
>
> "a doctrine of nature will contain so much science proper as there is applied mathematics in it" (Kant);
>
> "knowledge has become scientific to the extent that it is able to employ number and measure" (Nietzsche);
>
> "when you cannot measure ... your knowledge is ... meagre and unsatisfactory" (Thomson);
>
> "we hardly recognize a subject as scientific if measurement is not one of its tools" (Boring);
>
> "psychology advanced from prescience to science" "as a consequence of Fechner's [measurement] contribution" (Adler)
>
> "until the phenomena of any branch of knowledge have been submitted to measurement and number, it cannot assume the status and dignity of a science" (Galton);
>
> "psychology cannot attain the certainty and exactness of the physical sciences, unless it rests on a foundation of experiment and measurement" (JM Cattell);
>
> the "crucial method is that of measurement" (Spearman);
>
> "the universal validity of psychological results is guaranteed by the measurability of mental phenomena" (Kulpe);
>
> "science recognizes only quantities" (Freud); (all previous quotes from Michell, 1990, pp. 6-8);
>
> "in seeking the correct path to truth we should be concerned with nothing about which we cannot have a certainty equal to that of the demonstrations of arithmetic and geometry" (Descartes, 1961, p. 8);
>
> "the mathematical is the fundamental presupposition of the knowledge of things ... [and] of 'academic' work" (Heidegger, 1967, pp. 75-6);

> "the foundation of modern thought and knowledge is essentially mathematical" (Heidegger, 1967, p. 76);
>
> "it is *not word but number* that is the real paradigm of the noetic" (Gadamer, 1989, p. 412); and
>
> "the ideal [mathematical] object is the absolute model for any object whatsoever, for objects in general" (Derrida, 1989, p. 66).

Only recently, and not in all areas of the psychosocial sciences, has the automatic connection between mathematical thinking, quantification, and scientific status, referred to by Michell (Michell, 1990, 2000) as the quantitative imperative, been questioned. But too often the form this questioning takes implies a wholesale rejection of science, usually on the premise that quantification in the psychosocial fields is inherently reductionist, and that human behavior, attitudes, and abilities are better understood qualitatively (Martin and Sugarman, 2001).

A better-informed challenge to the quantitative imperative notes that standards for quantification and measurement are vastly different in the physical and psychological sciences. The idea that a field is scientific only to the extent that it is mathematical follows from the power wielded when researchers share a common quantitative language for meaningfully and usefully comparing amounts of a variable of interest. In psychology, the quest for so-called "gold standards" is doomed to failure as long as it depends on the definition of a single collection of test or survey items—a single brand of instrument—that all practitioners will agree on as meeting their needs. Because such agreement has been lacking, each researcher invents her or his own instrument(s), and each of these have different units of measurement, since they ask different numbers of questions, and observations are scored in different ways.

One way of stating the thesis of this article is that measurement will not make psychology scientific until each of the variables purportedly measured by the field's various instruments 1) are determined to be quantitative via experimental tests, and 2) are equated to a reference standard metric, so that everyone everywhere who is interested in a particular variable can participate in the ongoing conversation about that construct in a common quantitative language. The second of these points is the one that more obviously addresses interpretability, since a single reference standard range of quantities for each separate construct would be much more convenient and accessible than the current multiple ranges of alleged measures. The first point concerning experimental tests of quantitative structure is where interpretability comes most fully to bear, and will be the focus of this article.

“Reason Has Insight Only into that which It Produces after a Plan of its Own”

To test experimentally the hypothesis that a variable is quantitative (Michell, 1990, 2000; Wright, 1977; Fischer, 1989) is to ask the question of reducibility or irreducibility: does a set of observations exhibit enough of a uniform internal consistency to justify treating the summed score as a sufficient statistic? Sufficiency is what makes a score meaningful and is commonly assumed as a property of data that are used for comparing individuals’ or groups’ abilities, attitudes, etc. (Andersen, 1977). If, however, the summed score does not extract all available information from the observations, it fails to function as a sufficient statistic since something other than it is needed to reproduce the pattern of responses across the items for a respondent, or the pattern of responses across the respondents for an item. Sufficiency is important because it functions to support a rich qualitative sense of what makes measures interpretable.

The qualitative, interpretive critique of quantitative method remains incomplete until it encompasses an ancient and deeper qualitative sense of mathematical thinking. In the history of psychology, Brentano, a teacher of both Freud and Husserl, is recognized for appreciating the possible need for qualitative methods in psychology (Michell, 1990, p. 8). Husserl (Husserl, 1970; Derrida, 1989), in his “Origins of Geometry”, focused on what he called Galileo’s “fateful omission” of the means by which the mathematical understanding of nature arose. Husserl’s student, Heidegger (Heidegger, 1967), then picked up the theme and elaborated on Kant’s (Kant, 1929, p. 20) (B xiii) realization that, in the wake of the scientific revolution

> all students of nature...learned that reason has insight only into that which it produces after a plan of its own, and that it must not allow itself to be kept, as it were, in nature’s leading strings, but must itself show the way with principles of judgment based upon fixed laws, constraining nature to give answer to questions of reason’s own determining. Accidental observations, made in obedience to no previously thought-out plan, can never be made to yield a necessary law, which alone reason is concerned to discover.

This realization, “that reason has insight only into that which it produces after a plan of its own”, is crucial to connecting qualitative, interpretive methods with quantitative, experimental methods. The interpretation theories (hermeneutics) informing many contemporary psychosocial and cultural studies rightly take the mutual implication of subject and object, and so the social construction of knowledge, both in science and in society at large, as a basic supposition and theme. What is less widely appreciated is the role that this interaction and mutual implication of thing and thought plays in the mathematical metaphysics of science.

A wider sense of mathematics can begin from Descartes’ (Descartes, 1961, p. 17) own recognition that, for the ancient Greeks, “the name ‘mathematics’ means the same thing as learning.” This general sense of the Greek category ta mathemata as learning or as a curriculum of what can be taught and learned is widely recognized in the mainstream of the philosophy and history of mathematics (Bochner, 1966, pp. 24-26, 255; Bell, 1931, p. 58; Heath, 1931, p. 5; Dantzig, 1955, p. 25; Wilder, 1965, p. 284; Slavkov, 1973, p. 91; Miller, 1921, pp. 78, 17), but is rarely pursued at any length.

Heidegger, however, focused his attention on Greek mathematics in order to recover its wider implications for the conduct of science and the meaning of technological humanity (Heidegger, 1967, p. 67 ff; Heidegger, 1977a, pp. 115-154; Heidegger, 1977b, pp. 118-120). He says that

> *Ta mathemata* means for the Greeks that which man knows in advance in his observation of whatever is and in his intercourse with things: the corporeality of bodies, the vegetable character of plants, the animality of animals, the humanness of man. Alongside these, belonging also to that which is already-known, i.e., to the mathematical, are numbers. If we come upon three apples on the table, we recognize that there are three of them. But the number three, threeness, we already know. This means that number is something mathematical. Only because numbers represent, as it were, the most striking of always-already- knowns, and thus offer the most familiar instance of the mathematical, is ‘mathematical’ promptly reserved as a name for the numerical. In no way, however, is the essence of the mathematical defined by numberness. (Heidegger, 1977b, pp. 118-119; Heidegger, 1967, pp. 74-75)

In the age of modernity and classical physics, Copernicus, Galileo, Descartes and others felt so confident about the stability of the nature of knowledge and existence as things that would always be sure in their status as "always-already-knowns" that they were willing to define the essence of mathematics as only numerical; because of their situation in their particular historical context, perhaps they could not have thought or acted in any other way. But in the same way that the autonomous Cartesian subject and the priority of method over truth rely upon unexamined assumptions concerning the nature of knowledge and existence, so too does the definition of the mathematical as purely numerical.

So mathematical thinking has roots in both the interpretive construction of knowledge and in quantification and calculation (Fisher, 2003a, 2003b, 2004). This is important because each aspect serves to mitigate the shortcomings of the other. For instance, those who value quantitative precision criticize, in a manner reminscent of Plato, the various forms of constructivism as either mere rhetorical excursions that legitimize any assertion one cares to make, or solipsistic self-confirmations of expectations and prejudices. On the other hand, those who strive for a metaphysically-informed method (that accepts and incorporates realization of the fact that all understanding is constrained within the paradigms of language and culture) criticize psychology's quantitative methods for over-rigid categorization and unjustified assumptions of comparability that ignore, gloss over, or reductionistically homogenize differences that might be the most important or interesting features of a research population.

Contrary to what seems to be the predominant popular opinion in some areas of psychology, however, quantitative sciences, whether natural or human, are not inherently reductionistic or positivistic. One of the early innovators working both sides of the qualitative and quantitative mathematical divide remarked that "in scientific practice, ... the scientist often seems ... to be struggling with facts, trying to force them into conformity with a theory he does not doubt. Quantitative facts cease to seem simply 'the given.' They must be fought for and with, and in this fight the theory with which they are to be compared proves the most potent weapon" (Kuhn, 1961, p. 171; Kuhn, 1977, p. 193). Kuhn later adds that "some readers take my argument to mean that the committed scientist can make nature yield any measurements that he pleases. ... If what I have said is right, nature undoubtedly responds to the theoretical predispositions with which she is approached by the measuring scientist. But that is not to say either that nature will respond to any theory at all or that she will ever respond very much" (Kuhn, 1961, p. 176; Kuhn, 1977, p. 200). Heidegger (Heidegger, 1967, p. 93) concurs, saying that the realm of nature is axiomatically projected by mathematical expectations so that "a line of questioning can be instituted in such a way that it poses conditions in advance to which nature must answer in one way or another."

Gendlin (Gendlin, 1967, p. 292) remarks briefly that Kuhn's analysis of scientific revolutions are "highly consistent" with Heidegger's sense of the mathematical creation of meaning. Similarly, Gadamer (Gadamer, 1980, p. 164) says that Thomas Kuhn's elaborations on "the significance of the paradigm for the progress of research" do "not seem to me at all contradictory to the logic of inquiry." Both Heidegger and Kuhn focus on reason's own plans, and its capacities, however nonlinear, spiralling, or zigging and zagging these may be, to facilitate learning through those plans by projecting their structures in ways that allow the metaphoric process or the analogical imagination to have new experiences and to formulate new concepts by contrasting expectation and observation.

What keeps the process from becoming a vicious circle of self-fullfilling prophecy is the way in which scientific questioning remains open to unexpected answers, allowing experience to refuse the anticipated confirmation (Gadamer, 1980, p. 164; Gadamer, 1989, pp. 266, 350, 363, 464; Heidegger, 1962, p. 195). "Modern science is experimental because of the mathematical project" (Heidegger, 1967, p. 93), a project "where thinking thinks itself ..., i.e., [takes] cognizance of that which we already have" (Heidegger, 1967, p. 104). It is only by contrasting something recognizable as what is already known with something unexpected and unknown that new learning takes place.

This questioning attitude that connects known with known, performing the work that contextualizes and enframes the heretofore unexpected and unknown, permeates the practice of quantitative natural science, and has been self-consciously appropriated by interpretive, postmodern researchers in the psychosocial sciences and cultural studies. The question emerging at this point focuses on what a postmodern quantitative method of mathematical modeling would look like, especially in the simultaneous discovery and invention, or isolation and creation, of phenomena that can be consistently reproduced for study and measurement.

Symbols, Sufficient Reason, and Reduction

Reduction takes place as soon as reason forms and applies a plan of its own, even when that plan is no more fully articulated than unexamined linguistic concepts allow. The process by which the auditory and written symbols of language come to be recognized as standing for something is itself a form of reduction, an isolation of "this" from "all that" that occurs via interactions between the symbol, what it stands for, and those employing it (though it may actually be the case that the symbol does most of the employing). The relative isolation of some one kind of thing, a species or category, results in an indicator, a sound or a mark, that has come to delimit the conditions under which a particular phenomenon (variable, thing, construct) comes to a stand (Gadamer, 1989, p. 352). The symbol is not itself the phenomenon, but is all that is needed to evoke re-cognition of the phenomenon among those who understand it, i.e., those who have experienced the conditions under which the thing comes to a stand.

For instance, to say that I saw a brown dog in the street immediately evokes a variety of images belonging to the combined sets of brown colors, dogs, and streets. Provided with a color chart, you could pick out which shades are brown, which might be, and which are not. With the idea of a dog in mind, you could distinguish dogs from other animals, from plants, from machines, etc. Anyone trained in the language and who knows what a dog is, will, within some degree of error, recognize a dog as a dog.

What is important about this exercise is the sufficiency and invariance of the relationships between the symbols, the concepts, and the things themselves. The sounds and images that function as symbols for the concept of "dog" are sufficient and invariant indicators for a particular species of animal to the extent that they serve as consistent and reliable media for communicating and sharing meaning. That is, the kind of animal that counts as a dog typically exhibits a particular collection of features (fur, four legs, two eyes, two ears, a tail, paws, a mouth, anywhere from 2 to 100 pounds in weight, etc.) that are recognized together as sufficient for an indication that a particular animal is a dog. No matter what kind of dog is involved and no matter what speaker of the language in which "dog" is a symbol is involved, the abstract concept dog remains constant, retaining its structure independent of the samples of particular dogs and particular speakers involved in it at any one time.

The sufficiency of the name-concept relation in language provides the model for sufficient reductions (symbolizations) in mathematics, geometry, and scientific measurement. "Eudemos singles out Plato's contribution in his history of mathematics, namely, to have distinguished between name and concept (Simplicius *Physics* 98)" (Gadamer, 1980, p. 100).

Accordingly, "science and technology start from the universality of the concept" (Gadamer, 1989, p. 350). Plato placed philosophy in close association with mathematics because

> even he who has not yet seen all the metaphysical implications of the concept of pure thinking but only grasps something of mathematics . . . knows that in a manner of speaking one looks right through the drawn circle and keeps the pure thought of the circle in mind. (Gadamer, 1980, p. 101)

Gadamer could easily be paraphrasing Plato. In Book VI of the *Republic* (510d) Plato writes that mathematicians

> make use of the visible forms and talk about them, though they are not thinking of them, but of those things of which they are a likeness, pursuing their inquiry for the sake of the square as such and the diagonal as such, and not for the sake of the image of it which they draw.

Kant similarly points out that "the single figure which we draw is empirical, and yet it serves to express the concept, without impairing its universality" (Kant, 1929, p. 577) [A 714, B 742]. It is because he required students to comprehend that the object of philosophical inquiry is the abstract mathematical object that Plato put above the entrance to the Academy, "Let no one who has not grasped the mathematical enter here!" Accordingly, "the mathematical, in the original sense of learning what one already knows, is the fundamental presupposition of 'academic' work" (Heidegger, 1967, p. 76).

But in contrast to the contemporary sense of mathematical work as primarily a business of calculating, Plato is emphasizing the fact that numerical or geometrical figures and relationships are of less interest than the concepts they represent. It is important to remember that Plato was working in a Pythagorean world, since

> Pythagoreans take number and numerical relationships for existence itself and are unable to think of the noetic order of existence by itself, [so they never] see the real implications of the doctrine of ideas. (Gadamer, 1980, p. 35)

Even though the ancient Greeks had been faced with the crisis of the irrationality of some triangles' line segments throughout the century between Pythagoreas' death and Plato's writings (Dantzig, 1930, p. 26), "it is evident that prior to Plato a clear ontological conception of this difference [between name and concept] . . . was lacking" (Gadamer, 1980, p. 32); also see (Gadamer, 1989, p. 405). The condition of the possibility of science is the universality of the concept, and its paradoxical unity with and separation from the particular words and persons implicated in its birth, life, and death. As Gadamer (Gadamer, 1989, pp. 412-3) puts it,

> the sign has its being only in application, and so its 'self' consists only in pointing to something 'other'. It must be foregrounded from the context in which it is encountered and taken as a sign, in order for its own being as an object to be superseded and for it to dissolve (disappear) into its meaning.... The more univocally a sign-thing signifies, the more the sign is a pure sign—i.e., it is exhausted in the co-ordination. Thus for example, written signs are co-ordinated with particular sounds, numerical signs with particular numbers, and they are the most ideal signs because their position in the order completely exhausts them.

Numbers are the most easily recognized mathematical entities because they are the most ideal signs. The reason why science values number and measurement is because these are supposed to embody as nearly as possible the capacity for pure signification. In measurement and quantification, pure signification happens when an instrument most fully realizes the pointing function. When an instrument functions mathematically, it coordinates the question and answer process in such a way as to transparently bring the abstract, ideal thing itself into view. The instrument has to do this without imposing its own existence as an object on the object of inquiry. Instruments are instruments in the full meaning of the word only to the extent that they themselves are exhausted by the coordination of thing and thought, word and concept.

No instrument is ever fully transparent, and psychosocial measurement theory and practice needs to be especially attuned to the issues of error, reliability, and the internal consistency of data (Andersen, 1980; Andrich, 1988; Benson, 1998; Cherryholmes, 1988; Cronbach and Meehl, 1955; Cronbach, 1989; Messick, 1975; Messick, 1981; Messick, 1989; Wright, 1977; Wright and Masters, 1982). Even when error is acknowledged, so few instruments have been calibrated to be as transparent as possible, via tests of the quantitative hypothesis that check for the convergence and separation of name and concept, or word and thing, that much current psychological measurement remains "pre-mathematical" (Ballard, 1978, p. 187).

In prioritizing content validity over construct validity (Cherryholmes, 1988) in its tests and surveys, psychology focuses, in Pythagorean fashion, on the concrete line segments of individual figures, instead of looking through these at the abstract ideal. Because no effort is made to test the extent to which questions and answers belong together as consistently pointing together along a common line of inquiry, the numbers treated as measures remain dependent on the particular test or survey questions asked and on the particular sample responding. To the extent this is so, psychological researchers repeat the Pythagorean error of taking "number and numerical relationships for existence itself and are unable to think of the noetic order of existence by itself, [so they never] see the real implications of the doctrine of ideas" (Gadamer, 1980, p. 35).

Theoretical reasons why tests of sufficiency and invariance are important, and methods for performing the tests on psychological data, have been proposed in many forms over the last 75 years and more, in the works of Fisher (Fisher, 1922; Fisher, 1934), Thurstone (Thurstone, 1925; Thurstone, 1926; Thurstone, 1927; Thurstone, 1928), Guttman (Guttman, 1950), Loevinger (Loevinger, 1947; Loevinger, 1957), Luce (Luce, 1959; Luce, 1997; Luce and Tukey, 1964; Krantz, Luce, Suppes and Tversky, 1971; Suppes, Krantz, Luce and Tversky, 1989; Suppes, Luce, Krantz and Tversky, 1990), Rasch (Rasch, 1960; Rasch, 1961; Rasch, 1977), Coombs (Coombs, 1964), and others (Andrich, 1978c; Andrich, 1985; Brink, 1972; Brogden, 1977; Engelhard, 1984; Engelhard, 1994; Feinstein, 1987; Fischer, 1995; Michell, 1997a; Michell, 1997b; Perline, Wright and Wainer, 1979; van der Linden, 1994; Wright, 1985; Wright, 1999). When the quantitative hypothesis is not falsified, commonly used versions of these tests establish that the observed data (rating scale or test responses) support summarization as a sufficient statistic. Statistical sufficiency, identified by Ronald Fisher in the 1920s (Fisher, 1922; Fisher, 1934; Andersen, 1977; Fischer, 1995; Wright, 1980, pp. xi-xii; Wright, 1997), is one of several ways of establishing whether an ideal, abstract object has emerged from the question and answer interchange and taken on a life of its own, independent of the particular questions and answers that have given birth to it.

Hall, Wijsman, and Ghosh (1965) show "that the set of invariant rules based on a sufficient statistic is an essentially complete subclass of the class of invariant rules" (Arnold, 1988), where invariance is the stable structure and meaning of a quantitative unit across test/survey questions and respondents. One especially easy to apply approach to modeling sufficiency and invariance for different kinds of observations facilitated by tests and surveys emerged from the works of the Danish mathematician Georg Rasch. Rasch had studied with Fisher in London in the 1930s, and took special care to base his measurement models on the concept of sufficiency, stating a mathematical separability theorem that prescribes the data structures necessary for parameter invariance over samples of persons and items. Rasch focused on sufficiency because, as he (Wright, 1980, pp. xi-xii; Andrich, 1997) later said,

> When a sufficient estimate exists, it extracts every bit of knowledge about a specified feature of the situation made available by the data as formalized by the chosen model. 'Sufficient' stands for 'exhaustive' as regards the feature in question.
>
> What is left over when a sufficient estimate has been extracted from the data is independent of the trait in question and may therefore be used for a control of the model that does not depend on how the actual estimates happen to reproduce the original data....
>
> The realization of the concept of sufficiency, I think, is a substantial contribution to the theory of knowledge and the high mark of what Fisher did.... His formalization of sufficiency nails down the ... conditions that a model must fulfill in order for it to yield an objective basis for inference.

Fisher (1922) himself understood sufficiency as a crucial part of the mathematical foundations of theoretical statistics, deploring "the prolonged neglect into which the study of statistics, in its theoretical aspects, has fallen." Michell (1990, p. 130), echoing Guttman (1950), interpreted the neglect of the mathematical principles at the foundations of statistical inference as negligence of measurement theory, saying that

> In general psychologists have ... found refuge in quantitative methods that, because they assume more, demand less foundational research as the basis for their application. Methods that always yield a scaling solution, like the method of summated ratings, are almost universally preferred to methods which ... do not produce a scaling solution when they are falsified by the data. Surprisingly, vulnerability to falsification is commonly deemed by psychologists to be a fault rather than a virtue.

This is so even though the Encyclopedia of Statistical Sciences (Arnold, 1988, p. 79) states that "most statisticians accept the principle that statistical analysis should depend only on a sufficient statistic." Mining the same vein, Stevens (1951, p. 20) wrote that

> The scientist is usually looking for invariance whether he knows it or not.... The quest for invariant relations is essentially the aspiration toward generality, and in psychology, as in physics, the principles that have wide applications are those we prize.

Bachelard (Bachelard, 1984, pp. 36, 41), a philosopher of physical science, agrees, saying,

> it is in the determination of invariants that the mathematization of the real finds its true justification.... The whole problem of scientific knowledge of the real turns on the initial choice of mathematics. When one has fully comprehended ... that experimentation is always dependent on some prior intellectual construct, then it is obvious why one should look to the abstract for proof of the concrete.

It may be that tests of sufficiency and invariance are rare in the psychosocial sciences because most formulations of these tests have been so overly stringent that falsification of the quantitative hypothesis became a virtual certainty (Wilson, 1989).

Rasch (Rasch, 1960; Rasch, 1961; Rasch, 1977) formulated his approach to implementing and testing for sufficiency and invariance as a separability theorem, which, in contrast with deterministic measurement approaches, explicitly accepts that models are heuristic fictions. As Rasch (Rasch, 1960, pp. 37-8) put it, "models should not be true, but it is important that they are applicable....This also means that a model is never accepted finally, only on trial." The point is not, however, as some might think, to give up on obtaining sufficiency and invariance when data do not fit a model. Close examination of "failures of invariance" (Wimsatt, 1981) not only can illuminate data entry errors, ambiguous items, and simple mistakes, but can also lead to new insights into the structure of the variables of interest (Wright, 1977).

Be that as it may, mainstream quantitative methods are in much the same position now as they were in 1922 when Fisher decried the prolonged neglect of statistical theory. As Michell (Michell, 1990, p. 130) put it, "vulnerability to falsification is commonly deemed by psychologists to be a fault rather than a virtue," and very few measurement applications in the psychosocial sciences actually check for sufficiency and invariance, although theoreticians and philosophers have been showing for decades, if not centuries, that these are the necessary foundations of statistical inference, and are assumed to hold even when they are not demonstrated.

Constructing and Deconstructing Reductions: Modeling Laws

Software for tests and transformations is readily available (Adams and Kboo, 1995; Allerup and Sorber, 1977; ASC, 1996; Andrich, Lyne, and Sheridan, 2000; Andrich, Lynne and Sheridan, 2000; Fischer, 1997; Glas and Ellis, 1995; Gustafsson, 1979; Kelderman and Steen, 1988; Linacre, 1995; Smith, 1991; Linacre, 2004; Verhelst, 1993; Wu, Adams and Wilson, 1996), and the psychology and health care literature abounds with published examples of applications. A significant number of additional references are available when searching for related work on conjoint measurement, latent trait theory, item response theory, or validity studies.

Finally, "the reader who believes that all that is at stake in the axiomatic treatment of measurement is a possible canonizing of one scaling procedure at the expense of others is missing the point" (Ramsay, 1975, p. 262). The point is that we are faced with a choice between 1) investing in mathematically rigorous measurement, or 2) accepting that "it may not be possible to live forever with a dozen different procedures for quantifying the same piece of behavior, each making strong but untestable assumptions which result in nonlinear plots of one scale against another" (Ramsay, 1975, p. 262).

The axiomatic treatment of measurement is not a canonization of one scaling procedure at the expense of others because it does nothing to close the door 1) on the huge variety of methods by which systematic observations of behavior, attitudes, etc. can be made, 2) on the many different methods available for formulating testable quantitative hypotheses, or 3) on the many different ways of estimating parameters and assessing sufficiency, invariance, and consistency. The axiomatic projection of mathematical criteria for measurement quality is much less a canonization than it is what Gendlin (Gendlin, 1967, p. 291), in his commentary on Heidegger (1967), calls "a method of using many models, a method of using this human modeling power .... [that] is not this or that model, but the process of model-creating itself."

The entire Rasch family of mathematical models (Masters and Wright, 1984; Wright and Mok, 2000) implements his separability theorem in a manner that enables researchers to engage effectively in the process of model-creating itself for any measurable variable of interest. Each different model addresses the constraints of a particular observational framework. These frameworks might be structured by counts of correct answers from an examination (Rasch, 1960; Wright and Stone 1979), sums of ratings from a survey or performance assessment (Andersen, 1973; Andersen, 1977; Andrich, 1978a; Andrich, 1978b; Wright and Masters 1982), assignments of partial credit to multiple choice test answers (Andrich, 1978; Masters, 1982), ratings awarded by judges who vary in their propensity to be harsh or lenient (Linacre, 1989, 1997; Linacre, Engelhard, Tatum and Myford, 1994), or in a possibly infinite number of other ways (Adams and Wilson, 1996; Fischer, 1997; Wang, Wilson, and Adams, 1997, 1998; Wilson, 1989).

The point is that none of these inherently ordinal kinds of observations begin life as automatically reducible to a quantitative signifier-signified coordination. Whether or not the variable is quantitative must be modeled in a way that facilitates and tests for parameter convergence and separation. And there are many methods of varying complexity and exactness for implementing each model (estimating its parameters, error, and model fit): algebraic and approximative (Cohen, 1979), joint (unconditional) maximum likelihood (Wright and Douglas, 1977; Wright, 1988), pairwise comparisons (Choppin, 1968), fully conditional maximum likelihood (Andersen, 1973), and others, many of which are reviewed by Wright and Masters (1982), Linacre (1999), and Smith (2000).

Furthermore, the common usage that refers to these models as the dichotomous model, the rating scale model, the facets model, etc. is not entirely correct either. The point at which modelling actually takes place is the point at which specific sets of law-like regularities in a particular domain are addressed. The names of the parameters estimated, and the symbols for them, ought to be restated for each variable investigated as part of the process of formulating the quantitative hypothesis. The same variable might be addressed via several different multiple-choice, true-false, and short answer tests, as well as performance and portfolio assessments, for instance. If each different kind of observation is made on a separate sample, each data analysis *cum* instrument calibration will require a different model, but the names of the parameters would remain constant across models since they would hypothetically pertain to the same aspects of the same construct. If each different kind of observation, or some combination of them, is made on the same or linked samples, then all of the instruments could be equated onto a common metric in a single analysis that implements a mixed grouping of different submodels for each different kind of observation (Adams, Wu and Macaskill, 1997; Choppin, 1976; Engelhard and Osberg, 1983; Fisher, 1993, 1997a, 1997b, 1998, 1999, 2000; Fisher, et al., 1995, 1997; Gonin, Lloyd and Cella, 1996; Kelderman, 1986; Kolen and Brennan, 1995; Masters, 1985; Smith, 1996; Smith and Kramer, 1992; Wilson, 1994; Wright and Bell, 1984; Zhu, 1998).

Rather than one single "Rasch model," and rather than a family of generic models that apply to different kinds of data, what Rasch, measurement theory, and philosophy together provide are mathematical criteria for formulating and testing different models of specific constructs. If, for instance, the construct of interest to mathematics educators is hypothesized to be manifest in terms of correct responses to test questions, the model of interest states that mathematics ability b for person n minus the mathematical difficulty d of item i ought to be equal to the log-odds of success. The requirements here are 1) that any person with a particular score ought to have a higher probability of mathematics success on any item (whether or not it was on the test) than a person with a lower score, and 2) that any item with a particular score ought to have a higher probability of provoking success from any person (whether or not that person took the item) than an item with a lower score (Rasch 1960).

Because no test presents the entire universe of possible items to every possible examinee, meaningful inference from test scores requires generalization from the sample of persons who took the test to the population of all the people who might have, and from the sample of questions actually asked to the population of questions that might have been asked. Educators cannot purport to teach every conceivable aspect of a topic, or to test students on every conceivable problem that they might ever encounter. The nature of education and learning is such that we must presume to sample the universe of possible problems representatively enough to prepare students to deal with the world. What mathematically rigorous measurement enables us to do is to replace assumptions about tests and about the generalizability of measures with data quality requirements, based on principles of inference, that can be experimentally tested and improved for each individual construct.

Appendix: Reduction, Construction, and Creative Destruction

Reduction is one of three moments in the life cycle of knowledge about a thing, the other two being Construction and Creative Destruction (this latter is often referred to as deconstruction) (see Figures 1 and 2). Reduction (R in the two figures) involves abstraction and formalization, a qualitatively or quantitatively mathematical coordination of signifier and signified, with the goal of naming or generalizing some observation. In poetry, metaphor makes a new experience communicable via an interaction of its terms that opens up and conveys new meaning. All concepts are born as metaphors, and live only as long as they carry sufficient meaning to be constructively applicable. In science, mathematical models provide the most explicit and rigorous test of sufficiency and invariance in the form of instrument calibration and experimentation.

Reduction may be pure or applied. Even a poetic metaphor that conveys a successful reduction may never capture the imagination of an age and condense into a widely used concept. Similarly pure scientific research may entail logical consequences that have no practical application. In some cases, successful reductions may remain dormant for some time before being absorbed into an application.

Construction, the second phase in the life cycle of knowledge (C in the figures), is the application of a reduction in the building of community and communication. Communicability requires the common unity of meaning provided by reduction, though it rarely achieves the clarity it is often assumed to have. Constructive applications usually also entail instruction as a point of entry into the constructed world.

More specifically, then, construction is a community process of building together the structures of shared meanings that allow us to cooperate in sheltering, clothing, and feeding ourselves. These structures, based as they are in the reductions of sufficient reason, are mathematical in the ancient Greek sense of being teachable and learnable.

Far from conferring any kind of eternally self-evident status on the structures of reason, the fact that they are mathematical can do nothing to prevent the eventual *Destruction* (D in the figures) of reductions and the constructions built upon them. By the mere fact of their existence, constructions might alter their environments to the point of being unsustainable, entailing their self-destruction. Alternatively, critical thinkers might be able to detect the anomalies and inconsistencies in a reduction that make its constructions dysfunctional and/or meaningless, hopefully pointing the way toward improvements.

The deliberate cycling through the three phases of R, C, and D (as schematically drawn in Figure 1) in the life of knowledge and understanding constitutes a version of the hermeneutic circle, a circle that cannot be reduced to a vicious circle or to a tautology. It ought to be our "first, last, and constant task" to follow this circle through, to "keep the scientific theme secure", preventing our preconceptions and forestructures of knowledge from being presented to us only by unexamined prejudices and popular imagining (Heidegger, 1962, p. 195).

Thus the recent intense interest of academics in creative destruction need not be interpreted as a movement toward irrationality, as is sometimes the case (Gross and Levitt, 1994). It is, rather, but a moment in the broader evolution of our understanding of what knowledge is. Derrida (Caputo, 1997, pp. 49-70; Derrida, 1983), for instance, counters the criticism that his work is radically destructive by pointing to the new institutions he is involved in founding, institutions that necessarily require new reductions and constructions to structure them. He also stresses that the

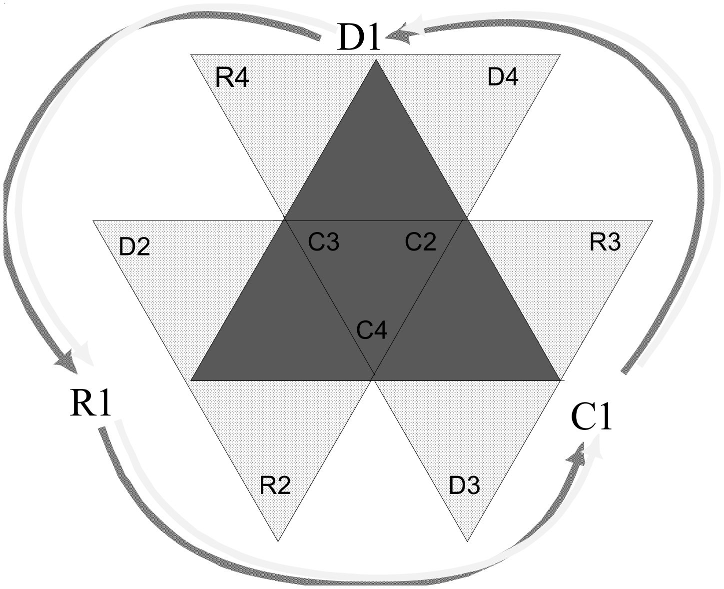

Figure 1. The highest level spiral in the ontological method: Reduction (R), Construction (C), and Creative Destruction/Deconstruction (D)

liberties he takes with texts are explicitly considered in relation to a distance measured from the standards that he was trained in (Derrida, 2003, p. 62). His focus from the beginning of his career was to place himself at the point at which "the thing signified is no longer easily separable from the signifier" (Derrida in Wood and Bernasconi, 1988, pp. 88-9).

One positive outcome of the recent wave of deconstructions could be a greater willingness among researchers to invest in the relatively intense foundational work required for deep quantitative understandings that reduce to valid models and metaphors. A second outcome of deconstruction could be a wider appreciation for the fact that all models and metaphors are abstractions that concrete data never match exactly. Any conceptual generalization necessarily entails error. When examined closely enough, any model fails, any reduction is strictly speaking reductionistic, because no data are perfectly sufficient or invariant. The point is to find out how much of a difference makes a difference.

Logical positivism was largely silent about reduction understood as sufficiency and/or invariance. It focused instead on a different, empirical kind of reduction, one that sought to simplify emotions, behavior, and cognition to biochemistry and physics. This form of reductionism has virtually disappeared from most areas of the psychosocial sciences, but has hardly even been challenged as a habit of mind in many areas of biology, chemistry, and physics.

Although psychosocial scientists who value a linguistic or cultural paradigm may feel threatened by classic positivist reductionism's hegemonic goals, they ought to feel far more threatened by the unexamined reductionisms operating within their own paradigms.

One of the major lessons that ought to be learned from deconstruction involves a deeper appreciation for the principle of sufficient reason (Heidegger, 1991), especially for tests that show these principles to be realized, as provisional and tentative as those realizations may be.

Most current psychosocial measurement practice is reductionistic in the sense of assuming sufficient reduction to be achieved, even when the treatment of raw scores (counts of right answers or summed ratings) as sufficient statistics is not even acknowledged, much less justified via articulation of a measurement model and tests of data, model fit.

Although Rasch (Rasch, 1960; Andersen, 1977) began from Fisher's (Fisher, 1922) notion of sufficiency in deriving his separability theorem, there are many approaches to measurement theory that provide satisfactory proofs of the theorem (Wright, 1997; Wright, 1999). What these approaches share is the focus on the raw score as all that is needed to reproduce the pattern of responses across items, for a person, and across persons, for an item.

Inference is always and everywhere a matter of understanding inherently unique events in terms that will be useful, meaningful, and reliable in managing the future. Even experiments in classical physics are not perfectly reproducible, meaning that concrete observations must be reduced to mathematically sufficient and invariant abstractions (laws or models) before they can be taken to predict future events. When observations are not mathematically evaluated for sufficiency and invariance, reduction and abstraction are not justified, and so are positivistically reductionist.

Psychological measurement is commonly both positivist and reductionist, in the sense of making broad and unsubstantiated assumptions concerning the extent to which discrete historical observations provide legitimate information concerning the structure of future general phenomena. Sufficiency and invariance are characteristic of scientific instruments, which historically are known for selectively filtering and shaping phenomena into stable reproducible forms for experimental study. The philosophical school of instrumental realists (Ackermann, 1985; Hacking, 1983; Heelan, 1983; Ihde, 1991; Bachelard, 1984) holds that scientific advances stem largely from the mathematical/dialectical interplay of instruments, data, and theory.

Unfortunately, as recently shown by historians of science, the role of instruments and the technicians who create and maintain them has been much neglected both in studies of science and in psychosocial measurement practice. This neglect, coupled with centuries, even millenia, of stern injunctions concerning the absolute necessity of mathematics to scientific thinking, has led to serious deficiencies in the quality of measurement in psychosocial studies. Psychological measurement practice could be less reductionist and positivist by placing greater value on explicit tests of sufficiency and invariance in the calibration of its instruments. Methods for conducting such tests date back to Thurstone's work in the 1920s, and have lately reached fairly widespread use in the application of Rasch's probabilistic formulations of axiomatic conjoint measurement theory.

When data are fit to a mathematical model requiring scores to function as sufficient statistics, and failures of invariance are actively investigated and rectified, reduction is justified. The test of model fit imply that the sought reduction^1 is provisionally put on hold, while all 3 phases of R^2-C^2-D^2 are cycled through at least once before moving on to construction^1 (see Figure 2). Construction^1 is then also temporarily put on hold while the instrument that has satisfactory calibrated in reduction^1 on pilot data is now checked for continued satisfactory performance in another checking through all three phases of R^3-C^3-D^3. Given quality functioning at phase C^1, standards are set, measures are reported, and decisions are made.

The paradigm—shifting consequences of a full D^1 analysis may occur less than once in a lifetime, century, or millenium, as may be happening with the increasingly critical appraisals of psychosocial measurement practices. A full D^1 analysis entails a fundamental reconceptualization of what reduction and construction themselves are, with far-reaching methodological consequences. Scientific revolutions, seen recently as involving new sociotechnical capacities to consistency reproduce standard phenomena (Latour, 1987, Fisher, 2000, 2004) probably always entail at

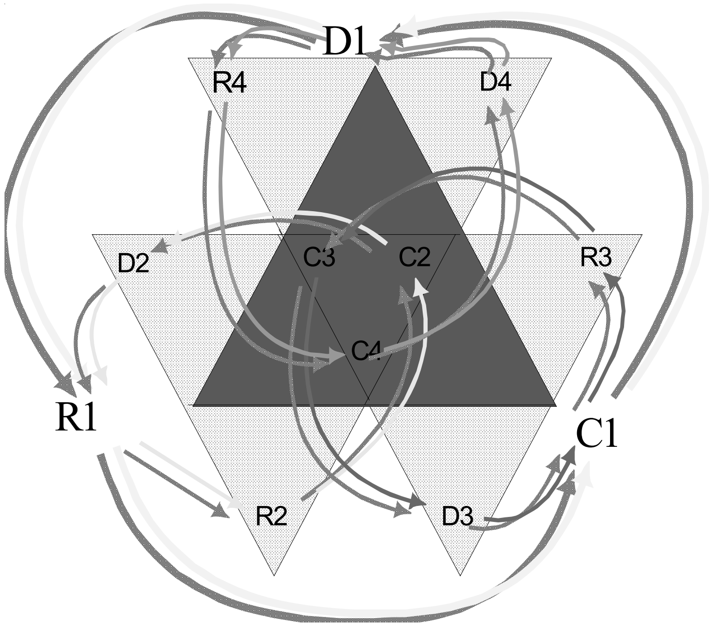

Figure 2. Lower level cycles across R, C, and D within each instantiation of R, C, and D.

Table 1

The Life Cycle of Facts

Phenomenology (method of ontology) (Heidegger 1927/1982: 21 ff.)	Poetry	History of mathematics	Music	Scientific method	Practice	Stage
Birth						
Reduction: “leading phenomenological vision back from the apprehension of a being ... to the understanding of the being of this being” (p. 21)	Tropes: metaphor, analogy, simile, synedoche, metonymy, etc.	Greek abstraction of number from things; i.e., 2+2=4 no matter what pairs of things are added	Scale/tonal organization; harmonics; instrument design and construction	Theory development; instrument calibration studies of invariance, sufficiency, additivity, concatenation, parameter separation, sample-free and scale-free measures, etc.	Experimentally tests basis for later construction and application by iterating through construction and deconstruction stages	Pre-paradigmatic: no agreement on reductive validity and no theory of variables, so no reference standards for quantity or quality
Life						
Construction: “projecting of the antecedently given being upon its being and the structures of its being” (p. 22)	Concepts and usage	Pythagorean theorem: axiomatic prescription of expected relationships; Euclidean geometry	Composition and performance within genre	Application of measures in research and practice	Experimental focus on evaluating effects of different treatments, aiming at a new level of reduction	Paradigmatic: reference standard quantities and QA in common use guiding practice and structuring communication, providing coherent sense of community
Death, Transformation, Rebirth						
Destruction: “a critical process in which the traditional concepts, which at first must necessarily be employed, are deconstructed down to the sources from which they were drawn” (p. 23).	Critical theory	Exploitation of anomalies leading to new reductions as Non-Euclidean, Hilbert, etc. geometries	Critical theory	Critical evaluation of measures, qualitative investigation of anomalies	Repeated application and instrument refinements isolate and highlight anomalies, exceptions that test the rule; struggle to explain these may provoke new paradigm for reduction	Post-paradigmatic: anomalies provoke widespread disagreement on legitimacy of current practice, provoking search for new summarization criteria

least an implicit full deconstruction of a pre-existing paradigm's R-C-D criteria for recognizing valid knowledge. Most research is conducted within an existing R-C-D paradigm, meaning that even truly innovative applications extending the paradigm into new fields attempt deconstruction only at the D^2 and D^3 levels. For research to cycle through D^1 it must achieve a paradigm shift that rearticulates the criteria for valid R^1 and C^1. Pure research remains within the R^1 (R^2-C^2-D^2) phase, never reaching application in the C^1 (R^3-C^3-D^3) phase. Applied research requires that a prior R^1 cycle(s) be completed, though it may never be explicitly related to the prior R^1 work.

Acknowledgments

Portions of this paper were presented at the 1999 American Psychological Association Annual Meeting in Boston, and at the Eighth International Objective Measurement Workshops held at the University of California, Berkeley, in 1995.

References

Ackermann, J. R. (1985). *Data, instruments, and theory: A dialectical approach to understanding science.* Princeton, NJ: Princeton University Press.

Adams, R. J., and Kboo, S.-T. (1995). *Quest: The interactive Rasch test analysis system, version 2.* Camberwell, Victoria, Australia: Australian Council for Educational Research. (Distributed in North America by Assessment Systems Corp., St. Paul, MN).

Adams, R. J., and Wilson, M. (1996). Formulating the Rasch model as a mixed coefficients multinomial logit. In M. Wilson (Ed.), *Objective measurement: Theory into practice, Vol. III* (pp. 143-66). Norwood, NJ: Ablex.

Adams, R. J., Wu, M. L., and Macaskill, G. (1997). Scaling methodology and procedures for the mathematics and science scales. In M. O. Martin and D. L. Kelly (Eds.), *Third international mathematics and science study technical report: Vol. 2: Implementation and analysis—Primary and middle school years*. Boston: Center for the Study of Testing, Evaluation, and Educational Policy.

Allerup, P., and Sorber, G. (1977). *The Rasch model for questionnaires. Report no. 16.* Copenhagen, Denmark: Danish Institute for Educational Research.

Andersen, E. B. (1977). Sufficient statistics and latent trait models. *Psychometrika, 42*(1), 69-81.

Andersen, E. B. (1980). *Discrete statistical models with social science applications.* Amsterdam: North-Holland.

Andersen, E. B. (1973). Conditional inference for multiple choice questionnaires. *British Journal of Mathematical and Statistical Psychology, 26,* 31-44.

Andersen, E. B. (1977). The logistic model for m answer categories. In W. E. Kempf and B. H. Repp (Eds.), *Mathematical models for social psychology*. Vienna, Austria: Hans Huber.

Andrich, D. (1978a). A binomial latent trait model for the study of Likert-style attitude questionnaires. *British Journal of Mathematical and Statistical Psychology, 31,* 84-98.

Andrich, D. (1978b). A rating formulation for ordered response categories. *Psychometrika, 43,* 357-374.

Andrich, D. (1978c). Relationships between the Thurstone and Rasch approaches to item scaling. *Applied Psychological Measurement*, *2*, 449-460.

Andrich, D. (1985). An elaboration of Guttman scaling with Rasch models for measurement. In N. B. Tuma (Ed.), *Sociological methodology 1985* (pp. 33-80). San Francisco: Jossey-Bass.

Andrich, D. (1988). *Rasch models for measurement.* Sage University Paper Series on Quantitative Applications in the Social Sciences, vol. series no. 07-068. Beverly Hills, CA: Sage Publications.

Andrich, D. A. (1997). Georg Rasch in his own words [excerpt from a 1979 interview]. *Rasch Measurement Transactions*, *11*(1), 542-3. [http://www.rasch.org/rmt/rmt111.htm#Georg].

Andrich, D. A., Lyne, A., Sheridan, B., and Luo, G. (2000). *RUMM: Rasch unidimensional models for measurement.* Perth, Australia: RUMM Laboratory Pty Ltd [www.rummlab.com.au].

Arnold, S. F. (1988). Sufficient statistics. In S. Kotz, N. L. Johnson, and C. B. Read (Eds.), *Encyclopedia of statistical sciences* (pp. 72-80). New York: John Wiley and Sons.

ASC. (1996). *RASCAL: Rasch analysis program.* St. Paul, MN: Assessment Systems Corp.

Bachelard, G. (1984). *The new scientific spirit* (Arthur Goldhammer and Foreword by Patrick A. Heelan, Trans.). Boston: Beacon Press.

Ballard, E. G. (1978). *Man and technology: Toward the measurement of a culture.* Pittsburgh, PA: Duquesne University Press.

Bell, E. T. (1931). *The queen of the sciences.* Baltimore: Williams and Wilkins.

Benson, J. (1998, Spring). Developing a strong program of construct validation: A test anxiety example. *Educational Measurement: Issues and Practice*, *17*(1), 10-17, 22.

Bochner, S. (1966). *The role of mathematics in the rise of science.* Princeton, NJ: Princeton University Press.

Bond, T., and Fox, C. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences.* Mahwah, NJ: Lawrence Erlbaum Associates [http://homes.jcu.edu.au/~edtgb/book/].

Brink, N. E. (1972). Rasch's logistic model vs the Guttman model. *Educational and Psychological Measurement*, *32*, 921-927.

Brogden, H. E. (1977). The Rasch model, the law of comparative judgment and additive conjoint measurement. *Psychometrika*, *42*, 631-634.

Caputo, J. D. (1997). *Deconstruction in a nutshell.* New York: Fordham University Press.

Cherryholmes, C. (1988). Construct validity and the discourses of research. *American Journal of Education*, *96*(3), 421-457.

Choppin, B. (1968). An item bank using sample-free calibration. *Nature*, *219*, 870-872.

Choppin, B. (1976). Recent developments in item banking. In D. N. M. DeGruitjer and L. J. van der Kamp (Eds.), *Advances in psychological and educational measurement* (pp. 233-245). New York: Wiley.

Cohen, L. (1979). Approximate expressions for parameter estimates in the Rasch model. *British Journal of Mathematical and Statistical Psychology*, *32*, 113-120.

Coombs, C. H. (1964). *A theory of data.* New York: Wiley.

Cronbach, L. (1989). Construct validation after thirty years. In R. Linn (Ed.), *Intelligence: Measurement, theory, and public policy: Proceedings of a symposium in honor of Lloyd Humphreys* (pp. 147-167). Urbana, IL: University of Illinois Press.

Cronbach, L. J., and Meehl, P. E. (1955, July). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281-301.

Dantzig, T. (1930). *Number, the language of science.* New York: Macmillan.

Dantzig, T. (1955). *The bequest of the Greeks.* New York: Macmillan.

Derrida, J. (1983). The principle of reason: The university in the eyes of its pupils. *Diacritics*, *13*, 3-20.

Derrida, J. (1989). *Edmund Husserl's origin of geometry: An introduction.* Lincoln, NE: University of Nebraska Press.

Derrida, J. (2003). Interview on writing. In G. A. Olson and L. Worsham (Eds.), *Critical intellectuals on writing* (pp. 61-9). Albany, NY: State University of New York Press.

Descartes, R. (1961). *Rules for the direction of the mind.* Indianapolis, IN: Bobbs-Merrill.

Engelhard, G., Jr. (1984). Thorndike, Thurstone, and Rasch: A comparison of their methods of scaling psychological and educational tests. *Applied Psychological Measurement*, *8*(1), 21-38.

Engelhard, G., Jr. (1994). Historical views of the concept of invariance in measurement theory. In M. Wilson (Ed.), *Objective measurement: Theory into practice*, *Vol. 2* (pp. 73-99). Norwood, NJ: Ablex.

Engelhard, G., Jr., and Osberg, D. (1983). Constructing a test network with the Rasch measurement model. *Applied Psychological Measurement*, *7*(3), 283-294.

Feinstein, A. (1987). *Clinimetrics.* New Haven, CT: Yale University Press.

Fischer, G. H. (1989). Applying the principles of specific objectivity and of generalizability to the measurement of change. *Psychometrika*, *52*(4), 565-587.

Fischer, G. H. (1995). Derivations of the Rasch model. In G. Fischer and I. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 15-38). Springer-Verlag.

Fischer, G. H. (1997). Structural Rasch models: Some theory, applications, and software. In M. Wilson, G. Engelhard, Jr., and K. Draney (Eds.), *Objective measurement: Theory into practice, Vol. 4* (pp. 185-208). Norwood, NJ: Ablex Publishing.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical transactions of the royal society of London, A, 222*, 309-368.

Fisher, R. A. (1934). Two new properties of mathematical likelihood. *Proceedings of the Royal Society*, *A*, *144*, 285-307.

Fisher, W. P., Jr. (1993). Scale-free measurement revisited. *Rasch Measurement Transactions*, *7*(1), 272-3.

Fisher, W. P., Jr. (1997a). Physical disability construct convergence across instruments: Towards a universal metric. *Journal of Outcome Measurement*, *1*(2), 87-113.

Fisher, W. P., Jr. (1997b, June). What scale-free measurement means to health outcomes research. *Physical Medicine and Rehabilitation State of the Art Reviews*, *11*(2), 357-373.

Fisher, W. P., Jr. (1998). A research program for accountable and patient-centered health status measures. *Journal of Outcome Measurement*, *2*(3), 222-239.

Fisher, W. P., Jr. (1999). Foundations for health status metrology: The stability of MOS SF-36 PF-10 calibrations across samples. *Journal of the Louisiana State Medical Society*, *151*(11), 566-578.

Fisher, W. P., Jr. (2000). Objectivity in psychosocial measurement: What, why, how. *Journal of Outcome Measurement*, *4*, 527-63.

Fisher, W. P., Jr. (2003, December). Mathematics, measurement, metaphor, metaphysics: Part I. Implications for method in postmodern science. *Theory and Psychology*, *13*(6), 753-90.

Fisher, W. P., Jr. (2003, December). Mathematics, measurement, metaphor, metaphysics: Part II. Accounting for Galileo's "fateful omission." *Theory and Psychology*, *13*(6), 791-828.

Fisher, W. P., Jr. (2004, October). Meaning and method in the social sciences. *Human Studies: A Journal for Philosophy and the Social Sciences*, *27*(4), in press.

Fisher, W. P., Jr., Eubanks, R. L., and Marier, R. L. (1997). Equating the MOS SF36 and the LSU HSI physical functioning scales. *Journal of Outcome Measurement*, *1*(4), 329-362.

Fisher, W. P., Jr., Harvey, R. F., and Kilgore, K. M. (1995). New developments in functional assessment: Probabilistic models for gold standards. *NeuroRehabilitation*, *5*(1), 3-25.

Gadamer, H.-G. (1980). *Dialogue and dialectic: Eight hermeneutical studies on Plato* (P. C. Smith, Trans.). New Haven, CT: Yale University Press.

Gadamer, H.-G. (1989). *Truth and method* (Translation revised by Joel Weinsheimer and Donald G Marshall, Trans.) (2nd ed. rev.). New York: Crossroad.

Gendlin, E. T. (1967). Analysis (W. B. J. Barton, and V. Deutsch, Trans.). In *What is a Thing?* by Martin Heidegger (pp. 245-96). South Bend, IN: Regnery/Gateway.

Glas, C. A. W., and Ellis, J. L. (1995). RSP 1.0: Rasch scaling program [Computer program]. Arnhem, Netherlands: CITO.

Gonin, R., Lloyd, S. R., and Cella, D. F. (1996, February). Establishing equivalence between scaled measures of quality of life. *Quality of Life Research*, *5*(1), 20-26.

Gross, P. R., and Levitt, N. (1994). *Higher superstition: The academic left and its quarrels with science.* Baltimore: Johns Hopkins University Press.

Gustafsson, J.-E. (1979). *PML: A computer program for conditional estimation and testing in the Rasch model for dichotomous items. Report no. 85.* Göteborg, Sweden: Institute of Education, University of Göteborg.

Guttman, L. (1950). The basis for scalogram analysis. In S. A. Stouffer and et al. (Eds.), *Studies in social psychology in World War II. Vol. 4: Measurement and prediction* (pp. 60-90). New York: Wiley.

Hacking, I. (1983). *Representing and intervening: Introductory topics in the philosophy of natural science.* Cambridge, UK: Cambridge University Press.

Hall, W. J., Wijsman, R. A., and Ghosh, J. K. (1965). The relationship between sufficiency and invariance with applications in sequential analysis. *Annals of Mathematical Statistics*, *36*, 575-614.

Heath, T. L. (1931). *A manual of Greek mathematics.* Oxford, UK: Oxford at the Clarendon Press.

Heelan, P. (1983, September). Perception as a hermeneutical act. *Review of Metaphysics*, *37*, 61-75.

Heidegger, M. (1962). *Being and time.* New York: Harper and Row.

Heidegger, M. (1967). *What is a thing?* (W. B. Barton, Jr. and V. Deutsch, Trans.). South Bend, IN: Regnery/Gateway.

Heidegger, M. (1977a). Modern science, metaphysics, and mathematics. In D. F. Krell (Ed.), *Basic writings* (pp. 243-282). New York: Harper and Row.

Heidegger, M. (1977b). The question concerning technology. In D. F. Krell (Ed.), *Basic writings* (pp. 283-317). New York: Harper and Row.

Heidegger, M. (1991). *The principle of reason* (R. Lilly, Trans.) (Original work published 1957). Bloomington, IN: Indiana University Press.

Husserl, E. (1970). *The crisis of European sciences and transcendental phenomenology.* Evanston, IL: Northwestern University Press.

Ihde, D. (1991). *Instrumental realism: The interface between philosophy of science and philosophy of technology.* The Indiana Series in the Philosophy of Technology. Bloomington, IN: Indiana University Press.

Kant, I. (1929). *Critique of pure reason* (N. K. Smith, Trans.) (Unabridged). New York: St. Martin's Press.

Kelderman, H. (1986). *Common item equating using the log-linear Rasch model.* Twente, Netherlands: Department of Education, University of Twente.

Kelderman, H., and Steen, R. (1988). *LOGIMO computer program for log-linear item response theory modelling.* Twente, Netherlands: Department of Education, University of Twente.

Kolen, M. J., and Brennan, R. L. (1995). *Test equating.* New York: Springer.

Krantz, D. H., Luce, R. D., Suppes, P., and Tversky, A. (1971). *Foundations of measurement. Volume 1: Additive and polynomial representations.* New York: Academic Press.

Kuhn, T. S. (1961). The function of measurement in modern physical science. *Isis*, *52*(168), 161-193.

Kuhn, T. S. (1977). *The essential tension: Selected studies in scientific tradition and change.* Chicago: University of Chicago Press.

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society.* New York: Cambridge University Press.

Linacre, J. M. (1989). *Many-facet Rasch measurement.* Chicago: MESA Press.

Linacre, J. M. (1995). FACETS Rasch analysis [Computer program]. Chicago: MESA Press.

Linacre, J. M. (1996). Generalizability theory and many-facet Rasch measurement. In G. Engelhard, Jr. and M. Wilson (Eds.), *Objective measurement: Theory into practice*, *Vol. 3* (pp. 85-98). Norwood, NJ: Ablex.

Linacre, J. M. (1999). Understanding Rasch measurement: Estimation methods for Rasch measures. *Journal of Outcome Measurement*, *3*(4), 382-405.

Linacre, J. M., Engelhard, G., Tatum, D. S., and Myford, C. M. (1994). Measurement with judges: Many-faceted conjoint measurement. *International Journal of Educational Research*, *21*(6), 569-577.

Loevinger, J. (1947). A systematic approach to the construction and evaluation of tests of ability. *Psychological Monographs*, *61*(4:#285).

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, *3*, 635-694.

Luce, R. D. (1959). *Individual choice behavior.* New York: Wiley.

Luce, R. D. (1997). Quantification and symmetry: Commentary on Michell, quantitive science and the definition of measurement in psychology. *British Journal of Psychology*, *88*, 395-398.

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

Martin, J., and Sugarman, J. (2001, April). Interpreting human kinds: Beginnings of a hermeneutic psychology. *Theory and Psychology*, *11*(2), 193-207.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*, 149-174.

Masters, G. N. (1985, March). Common-person equating with the Rasch model. *Applied Psychological Measurement*, *9*(1), 73-82.

Masters, G. N., and Wright, B. D. (1984, Dec). The essential process in a family of measurement models. *Psychometrika.*, *49*(4), 529-544.

Messick, S. (1975, October). The standard problem: Meaning and values in measurement and evaluation. *American Psychologist*, *30*, 955-966.

Messick, S. (1981). Constructs and their vicissitudes in educational and psychological measurement. *Psychological Bulletin*, *89*, 575-588.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement*. New York: Macmillan.

Michell, J. (1990). *An Introduction to the logic of psychological measurement.* Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Michell, J. (1997b). Reply to Kline, Laming, Lovie, Luce, and Morgan. *British Journal of Psychology*, *88*, 401-406.

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

Miller, G. A. (1921). *Historical introduction to the mathematical literature.* New York: Macmillan.

Perline, R., Wright, B. D., and Wainer, H. (1979). The Rasch model as additive conjoint measurement. *Applied Psychological Measurement*, *3*(2), 237-255.

Ramsay, J. O. (1975). Review of *Foundations of Measurement*, *Vol. 1*, by D. H. Krantz et al. *Psychometrika*, *40*, 257-262.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Danish Institute for Educational Research. (Expanded edition, 1980. Chicago: University of Chicago Press.)

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability* (pp. 321-333). Berkeley, CA: University of California Press.

Rasch, G. (1977). On specific objectivity: An attempt at formalizing the request for generality and validity of scientific statements. *Danish Yearbook of Philosophy*, *14*, 58-94.

Slavkov, S. (1973). Mathematics, knowledge, technological advancement. In *Science, Technology, Man.* Sofia: Bulgarian Academy of Sciences.

Smith, R. M. (1991). *IPARM: Item and person analysis with the Rasch model.* Chicago: MESA Press.

Smith, R. M. (1996). Item component equating. In M. Wilson (Ed.), *Objective measurement: Theory into practice*, *Vol. 3* (pp. 289-308). Norwood, NJ: Ablex.

Smith, R. M. (2000). Fit analysis in latent trait measurement models. *Journal of Applied Measurement*, *1*(2), 199-218.

Smith, R. M., and Kramer, G. A. (1992). A comparison of two methods of test equating in the Rasch model. *Educational and Psychological Measurement*, *52*, 835-846.

Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1-49). New York: John Wiley and Sons.

Suppes, P., Krantz, D., Luce, R., and Tversky, A. (1989). *Foundations of measurement, Volume II: Geometric and probabilistic representations.* New York: Academic Press.

Suppes, P., Luce, R., Krantz, D., and Tversky, A. (1990). *Foundations of measurement, Volume III: Representation, axiomatization, and invariance.* San Diego: Academic Press.

Thurstone, L. L. (1925). A method of scaling psychological and educational tests. *Journal of Educational Psychology*, *16*, 433-451.

Thurstone, L. L. (1926). The scoring of individual performance. *Journal of Educational Psychology*, *17*, 446-457.

Thurstone, L. L. (1927). A law of comparative judgement. *Psychological Review*, *34*, 273-286.

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology, XXXIII,* 529-544*The Measurement of Values*. Midway Reprint Series. Chicago: University of Chicago Press, 1959, pp. 215-233.

van der Linden, W. (1994). Fundamental measurement and the fundamentals of Rasch measurement. In M. Wilson (Ed.), *Objective measurement: Theory into practice* (pp. 3-24). Norwood, NJ: Ablex.

Verhelst, N. D. (1993). *One parameter logistic model computer program.* Arnhem, Netherlands: Cito.

Wang, W.-C., Wilson, M., and Adams, R. J. (1997). Rasch models for multidimensionality between and within items. In M. Wilson, G. Engelhard and K. Draney (Eds.), *Objective measurement: Theory into practice, Vol. 4* (pp. 139-56). Norwood, NJ: Ablex.

Wang, W.-C., Wilson, M., and Adams, R. J. (1998). Measuring individual differences in change with multidimensional Rasch models. *Journal of Outcome Measurement*, *2*(3), 240-65.

Wilder, R. (1965). *Introduction to the foundations of mathematics.* New York: Wiley.

Wilson, M. (1989). A comparison of deterministic and probabilistic approaches to learning structures. *Australian Journal of Education*, *33*(2), 127-140.

Wilson, M. (1989). Saltus: A psychometric model of discontinuity in cognitive development. *Psychological Bulletin*, *105*(2), 276-289.

Wilson, M. (1994). Comparing attitude across different cultures: Two quantitative approaches to construct validity. In M. Wilson (Ed.), *Objective measurement: Theory into practice, Volume 2* (pp. 271-294). Norwood, NJ: Ablex.

Wimsatt, W. C. (1981). Robustness, reliability and overdetermination. In M. B. Brewer and B. E. Collins (Eds.), *Scientific inquiry and the social sciences*. San Francisco: Jossey-Bass.

Wood, D., and Bernasconi, R. (1988). *Derrida and differance.* Evanston, IL: Northwestern University Press.

Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement*, *14*(2), 97-116.

Wright, B. D. (1980). Foreword, Afterword. In *Probabilistic models for some intelligence and attainment tests*, by Georg Rasch [Reprint; original work published in 1960 by the Danish Institute for Educational Research]. Chicago: University of Chicago Press.

Wright, B. D. (1985). Additivity in psychological measurement. In E. Roskam (Ed.), *Measurement and personality assessment.* North Holland: Elsevier Science Ltd.

Wright, B. D. (1988, Sep). The efficacy of unconditional maximum likelihood bias correction: Comment on Jansen, Van den Wollenberg, and Wierda. *Applied Psychological Measurement.*, *12*(3), 315-318.

Wright, B. D. (1997, Winter). A history of social science measurement. *Educational Measurement: Issues and Practice*, *16*(4), 33-45, 52.

Wright, B. D. (1999). Fundamental measurement for psychology. In S. E. Embretson and S. L. Hershberger (Eds.), *The new rules of measurement: What every educator and psychologist should know*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Wright, B. D., and Bell, S. R. (1984). Item banks: What, why, how. *Journal of Educational Measurement*, *21*(4), 331-345.

Wright, B. D., and Douglas, G. A. (1977). Conditional versus unconditional procedures for sample-free item analysis. *Educational and Psychological Measurement*, *37*, 47-60.

Wright, B. D., and Linacre, J. M. (1997). *A user's guide to BIGSTEPS Rasch-model computer program*, *v. 2.7.* Chicago: MESA Press.

Wright, B. D., and Masters, G. N. (1982). *Rating scale analysis.* Chicago: MESA Press.

Wright, B. D., and Mok, M. (2000). Understanding Rasch measurement: Rasch models overview. *Journal of Applied Measurement*, *1*(1), 83-106.

Wu, M. L., Adams, R. J., and Wilson, M. R. (1996). *MATS: Multi-aspect test software.* Camberwell, Victoria, Australia: Australian Council for Educational Research.

Zhu, W. (1998). Test equating: What, why, how? *Research Quarterly for Exercise and Sport*, *69*(1), 11-23.