Mathematics, Measurement, Metaphor and Metaphysics II

Accounting for Galileo's 'Fateful Omission'

William P. Fisher, Jr.
METAMETRICS, INC.

ABSTRACT. This paper continues the exposition begun in the previous paper (Fisher, 2003b) concerning philosophy's metaphysical insistence on rigorous figure–meaning independence, and its own distrust of that insistence, turning now to the potential for a new metrological culture that values both the full integration of mathematics and measurement, and frequent, vigorous challenges to that integration. Recent criticisms of psychological measurement as subject to a quantitative or methodological imperative are evaluated in terms of the history of academic metaphysics developed in the previous paper. The thesis is proposed and defended that quantitative instruments effectively embody hermeneutic-mathematical metaphysics' coordination of signifier and signified only when both within- and between-laboratory metrology studies are completed. Experimental tests of instrument functioning and social networks of laboratories collaborating in the creation and maintenance of metric standards are seen as vital to the emergence of a new metrological culture in the human sciences.

KEY WORDS: mathematical thinking, measurement, metaphysics, method, postmodernism

Quantification and Scientific Status

As long as the sciences dealt with variables that were relatively concrete, with physically observable effects, such as length, weight and electrical resistance, units could be concatenated (effectively laid end-to-end, so to speak). Concatenation effectively led, in a more or less obvious way, to a unified coordination of the properties of numbers with unit amounts of the thing measured. When the coordination did not occur, the mathematical project's anticipated confirmation was effectively refused by the irreproducibility of the results across experiments.

Researchers interested in measuring physically additive variables nonetheless found the craft of instrument calibration to be an exacting task. The theoretical and practical difficulties encountered in early efforts at measuring

THEORY & PSYCHOLOGY Copyright © 2003 SAGE Publications. VOL. 13(6): 791–828
[0959-3543(200312)13:6;791–828;039291]

physical properties, as documented by Heilbron (1993), Kuhn (1977) and Roche (1998), do not seem to have been markedly less daunting than the difficulties now encountered in measuring abilities, attitudes and behaviors (Fisher, 2001). Measurement theory for physically concatenatable variables is often considered to have culminated in the work of N.R. Campbell (1920), though Michell (1994) argues that more sophisticated approaches were developed in the less well-known works of Hölder and others.

Problems arose with modern thinking's negligence of its wider mathematical metaphysics (Fisher, 2003a, 2003b) as science provoked increasingly difficult questions of meaning and existence, but was unable to provide satisfactory answers to them (Husserl, 1970, p. 5). Additional problems arose when fields without obvious and concrete quantitative variables sought to be taken as serious sciences in the context of Galileo's 'fateful omission' (Husserl, 1970, p. 49), his failure to 'inquire back into the original meaning-giving achievement which . . . resulted in the geometrical ideal constructions'. Michell (1990, 1999) shows that human abilities, attitudes and performances came to be seen by some as intractable enigmas because of the difficulties experienced in their quantification, difficulties that might be overcome if the means by which nature came to be understood mathematically could be recovered and adapted to the human sciences. But, Michell contends, the problems encountered in the quantification of psychosocial attributes seemed so insurmountable that most researchers in the human sciences now purport to quantify variables more through a redefinition of what measurement is than through explicit tests of the hypothesis of an additive structure.

The previous paper in this series (Fisher, 2003b) traces the history of the academy's hermeneutic-mathematical metaphysics through the work of Husserl's students and philosophical descendants (Heidegger, Gadamer, Ricoeur and Derrida). This paper attempts to further advance toward a recovery of Galileo's 'fateful omission' by (1) relating Michell's sense of psychology as subject to a 'quantitative imperative' with the hermeneutic-mathematical metaphysics of academia; (2) considering the role of uniform value both in economics generally as well as in research economics; and (3) relating the mathematical project's requirement of universal uniform measures to the metrological processes of instrument calibration.

The Quantitative Imperative

Michell (1990; also see 1997a, 1997b, 1999, 2000) offers an incisive and well-documented critique of the scientific status of psychological measurement. The dialogue that he (Michell, 1997a, 1997b) provoked in the *British Journal of Psychology* with several prominent experts (Kline, Laming, Lovie, Luce, & Morgan, 1997) in psychological measurement was a landmark airing of crucial issues that had previously received little attention.

His analysis brings the issues raised in the present essay in direct contact with the field experiencing the struggle that may be most decisive in shaping the outcome of the culture wars. Michell (1990) remarks:

> 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. (p. 130)

> The only way to decide whether or not the variables studied in any particular science are quantitative is to put that hypothesis to the test. This essential step is missing in the development of modern psychology. (p. 8)

> [In the 1930s] some serious criticisms were made of [psychology's] supposed methods of measurement and this, by implication, raised questions about the entire enterprise of psychological measurement. The impact on psychology was catastrophic. Psychologists responded by redefining the concept of measurement, in the end accepting a definition so inflated as to rule out none of their methods. That this devalued the concept was ignored, for at least it gave the appearance of conformity to the quantitative imperative. (Michell, 1990, p. 9)

Apparently little has changed since Suppes and Zinnes (1963) observed that:

> All too often in the behavioral sciences a direct reading instrument is available (and used) despite the fact that its readings are not justified; the readings do not correspond to any *known* fundamental or derived numerical assignment. (p. 21)

Michell obviously intuits the value of hermeneutic-mathematical metaphysics (Fisher, 2003a, 2003b) in his sense of quantification as requiring experimental tests of the hypothesis that the variable of interest is in fact quantitative, that is, that the variable is divisible into the additive magnitudes necessary for meaningful numeric representations and that can be accomplished only by means of the relevant sign–thing coordinations. The mainstream practice of psychological measurement just as obviously and thoroughly participates in the positivist program's too-quick narrowing of the hermeneutic-mathematical spectrum to the numerical.

Michell effectively substantiates his claims by providing: (1) an historical perspective on the ubiquitous assumption that a field must employ quantitative methods to be scientific (the quantitative imperative); (2) support for the idea that qualitative methods, data and results can be just as scientific as or more scientific than quantitative ones; and (3) an introduction to fundamental measurement theory's criteria for testing the hypothesis that a variable is quantitative.

Additional support for Michell's sense of the quantitative imperative follows from its similarity with what Danziger (1985) called the 'methodological imperative' in psychology. Both Michell and Danziger argue against letting the structure of the number system be the sole structure to which theories must be accommodated, and both point out that quantitative methodology's commonly accepted status as the *sine qua non* of scientific conduct heightens the danger of mere unproductive rule-following. Both also argue in favor of qualitative approaches in order to reduce the constant, unrelenting pressure to force theories into the structure of only one kind of methodology. Neither, however, relates the imperatives of quantitative method to the metaphysics of academic thought, and so both are prevented from tracing the course of hermeneutic-mathematical thinking through the history of psychology as effectively as they otherwise might have.

Danziger's and Michell's point that the use of number is far from enough to make a field scientific has a long history of being respected within at least a small part of the field of psychological measurement, and within the history of philosophy as metaphysics. Neither author makes as much use of this history as one might expect, given the stress they place on the extent to which psychology and the human sciences have blindly submitted to unexamined assumptions concerning the value of expressing variables numerically. For instance, to take a readily available example, in his 1936 presidential address to the Psychometric Society, L.L. Thurstone (1937/1959b) remarked, 'a study can be quantitative without being mathematical' (p. 9). Thurstone also makes the point that a study can be mathematical without being quantitative, as is in fact the case with academia's mathematical metaphysics.

Thurstone elaborates by saying that 'merely to count noses or the answers in a test or seconds of reaction time or volume of secretion does not make a study either mathematical or scientific', and that 'this is not unlike the confusion by which arithmetical labor is sometimes called mathematical'. The distinction between the merely numeric or arithmetical and the mathematical is fundamentally one of experiment: 'modern science is experimental because of the mathematical project' (Heidegger, 1967, p. 93); that is, science is experimental because of the way the mathematical project's anticipation of particular answers to research questions remains open to contradiction and refutation. The mathematician Nicolas Bourbaki (1971, pp. 26, 35) acknowledges that both experiment and mathematics axiomatically project prior beliefs in structural forms, but is at a loss to explain how the 'most unexpected' close connections between experimental phenomena and mathematical structures could come about.

Explanations for these connections may emerge in time in a philosophical environment more fully cognizant of hermeneutic-mathematical metaphysics' role in experimental science. Thurstone (1937/1959b), for instance, relates mathematical thinking to experiment in two ways, one qualitative and the other quantitative. First, he observes, '[s]ome students who are themselves unable to develop a mathematical idea are nevertheless well able to comprehend an essentially mathematical formulation of a psychological problem with its implications and experimental possibilities' (p. 10). This qualitative appreciation for mathematical thinking as involving the experimental arrangements through which opportunities for new learning can be designed resonates with implications relevant to the gift character of hermeneutics (Fisher, 2003b; Kisiel, 1973). Second, Thurstone founded measurement practice and quantification in the 'crucial experimental test' of instrument functioning. This test continues to be of particular interest in measurement theory (Andrich, 1978; Engelhard, 1984; Tenenbaum, 1999; Wright, 1999), and appropriately so, since it is effectively the process through which the metaphysical assumption of figure–meaning independence is evaluated.

Michell's focus on tests of the hypothesis that a variable is quantitative is therefore situated quite squarely within a Thurstonian measurement paradigm and academia's hermeneutic-mathematical metaphysics. In particular, it would seem that his sense of psychology's blind submission to a 'quantitative imperative' (Michell, 1990) and affliction with a 'methodological thought disorder' (Michell, 1997a, 2000) would benefit from an analysis revealing how unexamined positivist habits of mind and associated behaviors subject us to counterproductive metaphysical assumptions. But Michell does not make explicit use either of Thurstone's concept of the 'crucial experimental test' or of the postmodern sense of the 'thesis of philosophy' to frame his arguments. Either would lend considerable support to his position with regard to the reasons both why psychological researchers adhere to insufficient methods and why fundamental measurement theory is of particular value. But before exploring these possibilities for supporting and extending Michell's analysis, another one suggested by Michell (1999) will be addressed.

Pathology sans *Etiology*

Michell (1999, pp. 213–216) revisits Cliff's (1992) five reasons why psychological researchers have not taken advantage of revolutionary work in measurement theory (citing primarily Krantz, Luce, Suppes, & Tversky, 1971; Luce, Krantz, Suppes, & Tversky, 1990; Luce & Tukey, 1964; and Suppes, Krantz, Luce, & Tversky, 1989). Cliff's five reasons are that the mathematics is foreign to most psychologists, that there are few striking examples of the advantages to be gained, that there are no guidelines for applying the theory to incomplete data, that the research style implied by fundamental measurement theory is unfamiliar to psychologists, and that psychologists were distracted by other interesting developments in their work. Michell concludes that Cliff's reasons for the failure of fundamental measurement theory are unsatisfactory, but offers only the pathology of a 'methodological thought disorder' as a possible explanation (Michell, 1997a, 2000) why the connection between quantification and scientific status should be both (1) so strong as to make insufficient methods the norm, and (2) too weak to make the advantages of sufficient methods compelling.

The path to a better explanation for the lack of fundamental measurement in the human sciences can be discerned in Michell's own work. The problem would seem to be one of meeting researchers in the human sciences on their own ground, in the terms of their existing frame of reference. This is in fact the point of deconstructing the history of metaphysics, as was undertaken in the prior paper in this series (Fisher, 2003b). But Michell opens up another avenue of approach. Not only does he stress experimental tests of additivity as necessary for meaningful quantification, he also criticizes reviews of developments in test theory for making no mention of fundamental measurement theory, 'despite attempts by Keats (1967), Brogden (1977), Perline *et al.* (1979) and Andrich (1988) to relate Rasch's theory to conjoint measurement' (Michell, 1999, pp. 212–213).

Some have long taken the relation between Rasch's theory and rigorous principles of fundamental measurement as established (Bond & Fox, 2001; Fisher & Wright, 1994; Rasch, 1960, pp. 109–125; Wright, 1984, 1985), and the principles of the connection have been further clarified in recent work (Karabatsos, 1998, 2001). Michell (1999) reveals his awareness of the connection between Rasch's models for measurement and fundamental measurement theory again when he says that, 'if Rasch's hypothesis is correct, the estimates can be regarded as measures of the ability involved' (p. 12). Michell does not use the word 'measures' loosely, as it would be in the absence of the relevant tests of what he calls either the quantitative hypothesis (Michell, 1990) or the measurability thesis (Michell, 1999). Elsewhere (Michell, 1999, p. 102) he seems to think that Rasch measurement theory interprets ordinal raw scores as measures, but this is false. The Rasch approach to testing for additivity takes advantage of the natural logarithm in the estimation of log-odds units, or logits, linearizing scale-dependent ordinal observations into generalized and invariant ratio/interval measures in much the same way that the key-dependent tunings of musical instruments are transformed into equal-tempered scales (Maor, 1994; Mathieu, 1997, p. 137; Sullivan, 1985).

Though he recognizes a connection between Rasch's theory and fundamental measurement theory, Michell (2000) ignores the many hundreds of publications presenting developments in Rasch theory and practice[1] when he says 'that the hypothesis upon which psychometrics stands, the hypothesis that some psychological attributes are quantitative, has never been critically tested' (p. 639). On the contrary, the large body of work in the areas of Rasch model fit statistics (see R.M. Smith, 2000, for a recent overview) and the evaluation of dimensionality using principal-components factor analyses of residuals (Linacre, 1998; E.V. Smith, 2002; R.M. Smith, 1996) are in fact explicitly focused on tests of conjoint additivity. The quantitative status of psychological attributes is routinely assessed in Rasch-based research (e.g. S.K. Campbell, Wright, & Linacre, 2002; Dawson, 2002; Fisher, 1999a; Gehlert & Chang, 1998). The hypothesis that psychological attributes are quantitative is also effectively tested when different instruments intended to measure the same variable are calibrated on separate samples of respondents in independent studies, but are nonetheless found to measure the same construct in linearly transformable versions of the same metric (Fisher, 1997a).

Had he followed through with his recognition of the place Rasch's theory holds in the measurement tradition, Michell could have provided a much different assessment of the five reasons given by Cliff for the measurement 'revolution that never happened'. In opposition to Cliff's contentions that fundamental measurement's mathematics are foreign to psychologists and that there are few striking examples of its utility, Rasch's models have been employed in research and high stakes testing for decades (e.g. Choppin, 1976; Connolly, Nachtman, & Pritchett, 1971; Jaeger, 1973; Kelley & Schumacher, 1984; Rentz & Bashaw, 1977; Woodcock, 1973; Wright & Bell, 1984). The problem of error (incomplete and fallible data) has been addressed at length, with individualized error and internal consistency (model fit) estimates routinely provided for each person measured and item measuring (R.M. Smith, 2000; Wright & Masters, 1982; Wright & Stone, 1979). The error problem has been solved to the extent that item banking (Choppin, 1968, 1976; Wolfe, 2000; Wright & Bell, 1984), adaptive test and survey administration (Bergstrom & Lunz, 1999; Lunz, Bergstrom, & Gershon, 1994; Reckase, 1989; Weiss, 1983; Wright & Douglas, 1975) and multifaceted models for judged ratings (Linacre, 1989, 1996; Linacre, Engelhard, Tatum, & Myford, 1994; Lunz & Linacre, 1998) take explicit advantage of the capacity of Rasch's models to account for missing data. Given the volume of work produced, the researchers employing Rasch models in their work do not seem to have been distracted by other developments in their field. Finally, however, there is little doubt that the research style required for successful application of a Rasch model is very foreign to most researchers, with marked disagreement as to its utility and theoretical value (Andrich, 1989a, 1989b, 2002; Fisher, 1994).

Despite the ongoing debate, the sweeping generalization that psychometricians have never critically tested the quantitative hypothesis nonetheless commits the same error that Michell (1999) himself points out: 'a science that ignores a body of theory which is relevant to the testing of some of its hypotheses is denying itself an opportunity' (p. 219). The opportunity here is threefold: (1) understanding how the positivist abhorrence of metaphysics led to an incomplete definition of measurement (Fisher, 2003b); (2) understanding how the deconstruction of metaphysics demythologizes number and opens onto a qualitatively richer understanding of quantity and mathematical thinking (Fisher, 2003b); and (3) understanding how the advantages of measurement built out of research implementing demanding tests of additivity, sufficiency and parameter separation make the trouble of the investment worthwhile. The latter opportunity is addressed next.

Mathematical Language: Agent and Product of Agreement

Political Economies of Uniform Value

The fact that counts of right answers or sums of ratings are not measures was recognized by the early educational measurement theoretician E.L. Thorndike (1904), who observed that:

> If one attempts to measure even so simple a thing as spelling, one is hampered by the fact that there exist no units in which to measure. One may arbitrarily make up a list of words and observe ability by the number spelled correctly. But if one examines such a list one is struck by the inequality of the units. All results based on the equality of any one word with any other are necessarily inaccurate. (p. 7)

This passage would seem to contradict Michell's (1999) opinion that 'Thorndike mistakenly thought that observed scores count units of variable magnitude' (p. 102), though Thorndike might later have arrived at such a conclusion. It is true, as Thorndike's contemporary Binet understood (Michell, 1999, p. 94), that we do not really want to know how many words from a list someone can spell correctly; we are actually more interested in an abstract sense of how able a speller a person might be no matter what specific word comes up, so that the amount of ability would be 'superposable' or transferable across spelling tests in the same way that an amount of length is transferable across rulers. Just as in grocery store purchasing decisions we are less interested in a concrete number of apples (the count) than we are in their abstract measure (weight), so, too, ought we to distinguish between concrete counts and abstract measures in the human sciences (Wright, 1999).

Apples, for instance, are priced by weight rather than by count because amounts of apple remain constant for a given weight more consistently than they do for a given count. To base bulk produce pricing solely on counts with no additional information concerning volume amount or mass would create opportunities for unscrupulous profiteering by those able to buy large apples for the same price that they sell small apples for. The economies of many areas of research are in precisely this position of lacking uniform common currencies and making do with sample-dependent counts of correct answers, sums of ratings, and frequencies based on them.

Fair trade has historically been promoted by the deployment of uniform weights, standards and currencies. The Bible, the Torah, the Koran, the Magna Carta and the US Constitution all explicitly address weights and measures in the context of fairness and justice (Klein, 1974, pp. 29–33, 90–93; Wright, 1999). Furthermore, metrologically enabled societies enjoy higher standards of living and broader cultural influence than non-metrological societies. For instance, Western Europe rose to world dominance between 1250 and 1600 in part because of the practical leverage that effective quantitative technologies brought to bear on navigation, shipbuilding and weapons manufacturing (Crosby, 1997).

Not surprisingly, then, fields requiring stronger and more rigorous forms of inference make more rapid gains than fields that accept weaker and more lax inferential standards (Platt, 1964). The economies of health care (Coye, 2001; Kindig, 1997) and education, for instance, are often observed to behave in ways quite uncharacteristic of other economies. It may turn out that these idiosyncratic behaviors stem from the quality of these fields' inferential standards. Quality assessment and improvement procedures in these fields, employing the quantitative methods of psychology so thoroughly criticized by Michell, make no investment in calibrating instrumentation by coordinating numerical figures with empirical amounts of the variable of interest. In other words, the economies of education and health care fail to transcend the vagaries of local politics because the objects of their conversations are expressed as concrete counts of units that vary in size and order to unknown degrees across each particular person and question involved in them.

Because the instruments purported to measure educational and health care outcomes are not equated into coordinated sign systems—so that a given number consistently represents a single amount of the variable across brands of instruments; schools, examinees and teachers; and clinics, hospitals, clinicians and patients, and so on—the economies of these fields lack common quantitative currencies capable of serving as the basis for exchanges of proportionate value. To put the matter in the terms developed by Latour (1987, p. 223) and De Soto (2000), human capital as currently measured is effectively dead capital, in the sense that the measures are not transferable, being expressed as they are in scale-dependent, non-linear metrics unrecognized and unaccepted outside the bounds of a locally restricted community of practice (Fisher, 2002b, 2003a).

Developing along analogous lines, Wright's (1999) sense of measurement as facilitating the application to future problems of what is learned from the past strongly echoes actor-oriented transfer theory's focus on the dynamic production of invariance across contexts in mathematics learning (Lobato, 2003; Lobato & Siebert, 2002). Considerable evidence that invariant reference standard metrics can be calibrated, and human capital can thereby be brought to life, has accumulated over the last forty years. Such metrics are implied in every successful application of a Rasch model (Fisher, 1997b, 2000a), of which there are by now hundreds or thousands (Bond & Fox, 2001; Fisher, 2002a; Fisher & Wright, 1994).

It is not just empty speculation to hope, then, that the emerging global economy may entail a variety of cultural formations that effectively address Ricoeur's sense of the task we face: 'All the struggles of decolonization and liberation are marked by the double necessity of entering into the global technical society and being rooted in the cultural past' (Ricoeur, 1974, p. 292; also see Ihde, 1990, pp. 124–161). Or, as Derrida (1978) addresses the issue, 'the quality and fecundity of a discourse are perhaps measured by the critical rigor with which this relation [of unavoidable ethnocentrism] to the history of metaphysics and to inherited concepts is thought' (p. 282). If it is possible to succeed in raising the epistemological question on ontological grounds (Ricoeur, 1981, pp. 88–89; also see Fisher, 2003b; Ricoeur, 1978, p. 156), if we can begin measuring and managing human, social and natural capital resources from within different cultural frameworks delimited by the metaphors of each natural language, then each separate culture will have its own unique point of entry into the global technical society opened up in its own terms.

In this context, Ricoeur's (1979, pp. 244–246) analysis of Carnap's (1959) positivist distinction between logical object-sentences and metaphysical pseudo-object-sentences implies that requiring the construction of the positivist's artificial language would delay, obstruct and possibly prevent entry into the global technical society for any culture unwilling to abandon its own roots. Should it be possible to formulate an approach to measurement structurally analogous to or harmonious with metaphor, however, then each separate culture would have the means to extend its own measure of itself from its roots into the canopy of the global economy. This is what is at stake in the construal of the metaphoric process as the virtual calibration of linguistic instruments undertaken elsewhere (Fisher, 1988).

In current practice, we mistakenly rely on concrete counts as much as we do in testing and survey research because of the extent to which we simply assume, in Pythagorean fashion, driven blindly by the mathematical metaphysics behind the quantitative imperative, that the book of nature is written in a quantitative language. Elucidation of the hermeneutic-mathematical continuum and the history of academic metaphysics provides an etiology of the pathology described by Michell (1997a, 2000) as psychology's 'methodological thought disorder'. Michell is correct to say that scientific status is not the exclusive province of quantitative fields of study, and that we need a new experimental attitude toward testing the quantitative hypothesis to make the human sciences a part of the broader mathematical project.

But Michell stops short of realizing that the missing compelling application, the one whose absence prevents researchers and institutions from making the needed investment in more rigorous measurement methods and principles, is metrological generalization, the calibration and dissemination of a metric that constitutes the living language of a field of investigation by means of the clarity and portability with which constant amounts are signified across applications, samples and brands of instruments. When a common language is mobilized within a network of shared signification, with its terms and symbols everywhere recognized and accepted by those trained in reading them, meaningful communication is achieved, shared understandings and histories are more easily accumulated, and collective productivity is markedly enhanced (Bazerman, 1994; Hutchins, 1995; Latour, 1995; Maasen & Weingart, 2001; O'Connell, 1993; Shapin, 1994; Wegner, 1991; Wise, 1995).

There is little doubt that we need, methodically and routinely, to ask when measurement is justified and when it is not, in order to realize the full hermeneutic-mathematical range of the language in which the book of nature, human and otherwise, is written. But the diagnosis of psychology's pathology advanced by Michell (as well as the similar diagnoses offered by Andrich, 1989b; Cliff, 1993; Guttman, 1985; Wilson, 1971; Wright, 1984; and others) does not take the etiology to the point at which effective treatments suggest themselves. The remainder of this paper attempts to provide the needed correction by taking up a hermeneutic-mathematical exposition of quantitative languages as transferable representations of intellectual capital mobilized in and via networks of actors sharing common interests and values.

The Crucial Experimental Test: Integrating Hermeneutics and Mathematics in a Practical Procedure

Mathematically and metaphysically astute instrument calibration has the special advantage of fostering shared quantitative languages: uniform numerical expressions of commonly studied variables that function as reference standard metrics for all brands of instruments purported to measure those variables, and for any individual person who belongs to the relevant population. Heidegger (1967) points out that the mathematical project itself entailed the emergence of such reference standard metrics in the natural sciences:

> Because the [mathematical] project establishes a uniformity of all bodies according to relations of space, time, and motion, it also makes possible and requires a universal uniform measure as an essential determinant of things, i.e., numerical measurement.
>
> The mathematical project of Newtonian bodies leads to the development of a certain 'mathematics' in the narrow sense. The new form of modern science did not arise because mathematics became an essential determinant. Rather, that mathematics, and a particular kind of mathematics, could come into play and had to come into play is a consequence of the mathematical project. (p. 93)

After all, if the goal of quantification is to take advantage of the way in which number is the most obvious of mathematical entities, the 'always-already-known', then is not the point of instrument calibration a matter of doing the experimental, 'co-agital' work that will support the generalization of measured amounts in a coordinated sign system? Should not instrument calibration be the processual embodiment of the metaphoric process? Should not the end result of the calibration process be a number line capable of functioning as a common currency for the exchange of consistent amounts of the thing of interest, that is, quantitative value?

Far from following solely from the mere assignment of numbers according to a rule and resulting in a collection of numbers amenable to arithmetical manipulations devoid of substantive meaning, the real value of measurement stems from what Roche (1998) refers to as 'the true union of mathematics and measurement' (p. 145), and what Kuhn (1977) calls 'the full and intimate quantification' of a science (p. 221). Following Thurstone (1937/1959b), our aim is to arrive at a point where 'we are dealing with an equation whose parameters have meaning in terms of the psychological postulates that the equation represents' (p. 5), so that we are in a better position to appreciate the fact that 'mathematics . . . function[s] not merely as an aid or tool that a psychologist can use but as the very language in which he thinks' (p. 10). To the extent that the mathematical project's metaphysics comes into play in the quantitative methods of the human sciences, the emergence of 'universal uniform measures' should be expected. Indeed, paralleling the development of Heidegger's thought from science as the mathematical project to science as technology (Glazebrook, 2000, pp. 64, 252), we should expect the hermeneutic-mathematical infrastructure of science to extend to its instruments (Heelan, 1983a, 1983c; Ihde, 1991, 1998), and from there to the social networks of trust (Shapin, 1994) through which the reference standards and conventions associated with the use of those instruments are shaped (Fisher, 1997b, 2000a, 2002b; Latour, 1987; O'Connell, 1993).

Instruments effectively embody the hermeneutic-mathematical metaphysics' coordination of signifier and signified only when both within- and between-laboratory experimental studies are completed. Measurement theory in psychology typically focuses on the within-laboratory experiments alone, holding that a variable is shown to be quantitative only if it exhibits specific properties of additivity that remain invariant across samples of persons and questions, and if procedures can be devised for estimating ratios of magnitudes relative to some additive relation (Krantz et al., 1971; Luce et al., 1990; Luce & Tukey, 1964; Michell, 1994, p. 400; Suppes et al., 1989). Rasch's probabilistic conjoint measurement models (Andersen, 1980; Andrich, 1988; Fischer, 1974; Rasch, 1960, 1961, 1966; Wright, 1968, 1977; Wright & Mok, 2000) are perhaps the most well-known and widely used models requiring sufficiency and additivity (Andersen, 1977; Brogden, 1977; Fischer, 1974, 1995; Karabatsos, 1998, 2001; Perline, Wright, & Wainer, 1979; Wright, 1968, 1977, 1985, 1999), providing a context for assessing data quality (Adams & Wright, 1994; Andersen, 1973; Glas & Verhelst, 1995; Klauer, 1995; Linacre, 1998; Rost & von Davier, 1994; E.V. Smith, 2002; R.M. Smith, 1985, 1996, 2000), for calibrating instruments (e.g. Bayley, 2000; Bond & Fox, 2001; S.K. Campbell et al., 2002; Carlson, Andrews, & Bickel, 1999; A.G. Fisher, 1993; Fisher & Wright, 1994; Granger & Gresham, 1993; Kalinowski, 1985; Lunz, Wright, & Linacre, 1990; McNamara, 1996; R.M. Smith, 1997; Wright & Stone, 1979), and for connecting instruments into common metrics via equating (Dawson, 2002; Fisher, Eubanks, & Marier, 1997; Fisher, Harvey, Taylor, Kilgore, & Kelly, 1995; Masters, 1984, 1985). In order for the full union of mathematics and measurement (Roche, 1998) to be achieved, the mathematically rigorous results of Rasch analysis need to be moved outside the laboratories in which the work is done and connected together for use at the point the information is most needed, in the classroom, clinic and workplace (Fisher, 2000a). To do this, in turn, greater attention will have to be paid, first, to the practical consequences of the overlap between the hermeneutic and mathematical (Fisher, 2003a; Kisiel, 1973), and, second, to the inter-laboratory studies through which different instruments are equated together in common, unified metrics.

Wright (1999) and Fischer (1995) document the historical contributions to measurement made by a diverse array of theoreticians, showing that these are effectively and efficiently encapsulated in Rasch's (1960, 1961) separability theorem, concept of specific objectivity (Rasch, 1977) and associated series of models (Fischer & Molenaar, 1995; Masters & Wright, 1984; Wright & Mok, 2000). The theme of universal uniform measures, common to a wide variety of well-documented but rarely applied measurement approaches, is summarized by Thurstone (1928/1959a):

> The scale must transcend the group measured—One crucial experimental test must be applied to our method of measuring attitudes before it can be accepted as valid. A measuring instrument must not be seriously affected in its measuring function by the object of measurement. To the extent that its measuring function is so affected, the validity of the instrument is impaired or limited. If a yardstick measured differently because of the fact that it was a rug, a picture, or a piece of paper that was being measured, then to that extent the trustworthiness of that yardstick as a measuring device would be impaired. Within the range of objects for which the measuring instrument is intended, its function must be independent of the object of measurement. . . . [Accordingly,] if the scale is to be regarded as valid, the scale values of the statements should not be affected by the opinions of the people who help to construct it. This may turn out to be a severe test in practice, but the scaling method must stand such a test before it can be accepted as being more than a description of the people who construct the scale. (p. 228)

Thurstone (1926) also addressed the converse issue, in which the scale values interpreted as measures of individual performance, ability or attitude should not be affected by the particular questions asked.

Rasch formalizes Thurstone's 'crucial experimental test' with his separability theorem:

> On the basis of [one of the equations in the model] we may estimate the item parameters independently of the personal parameters, the latter having been replaced by something observable, namely, by the individual total number of correct answers. Furthermore, on the basis of [the next equation] we may estimate the personal parameters without knowing the item parameters which have been replaced by the total number of correct answers per item. Finally, [the third equation] allows for checks on the model [another equation] which are independent of all the parameters, relying only on the observations. (Rasch, 1961, p. 325; see also 1960, p. 122)

Both Rasch and Thurstone focus on establishing conjoint question-and-answer relations mediated by instruments off which are read quantitative values that remain invariant across samples of respondents and questions. The relations that hold between Thurstone's and Rasch's approaches to measurement are well known (Andrich, 1978, 1989a; Brogden, 1977; Engelhard, 1984, 1994; Tenenbaum, 1999), most notable being their experimental approach to parameter estimation, tests of the hypothesis that the variable is quantitative, and the evaluation of invariance.

Of course, the hermeneutic-mathematical implications of Rasch's separability theorem vis-à-vis the rigorous independence of numeric figures relative to amounts of things are rarely noted (Fisher, 1992, 1994, 2003a, 2003b). Luce and Tukey (1964, p. 4) express the opinion that the equivalence of simultaneous conjoint measurement's kind of objectivity with the objectivity of measurement in classical physics is 'only' of philosophical interest, overlooking the fact that this interest aims to advance the general understanding of the efficacy of science, and the etiology of psychology's methodological thought disorder. Rasch recognized, however, that there is something fundamental common to the observational material of all sciences, saying that for science to

> require observations to be measurable quantities is a mistake, of course; even in physics observations may be qualitative. . . as in the last analysis they always are! (e.g. as reading off a point *as located* between two marks on a measuring rod). (Rasch, 1977, p. 68)

Rasch here touches on Heelan's (1983a, 1983b, 1983c) construal of measuring instruments in physics as readable technologies, opens the matter onto Heidegger's (1962, pp. 469, 464–479) explication of time as read from shadows and watches, that is, as texts to which all the implications of hermeneutics apply, and also raises the issues noted by Kisiel (1973) in his comparison of hermeneutics and mathematics. Although Rasch is not trained in philosophy and does not develop the theme of a hermeneutic-mathematical continuum of figure–meaning coordinations, he does recognize that

> . . . on a first sight the observational material in Humanities would seem very different from that in physics, chemistry and biology, not to speak of mathematics. But it might turn out that the difference is less essential than it would seem. (Rasch, 1977, p. 68)

To demonstrate that the differences between the observational materials of the natural and human sciences are less essential than commonly assumed, Rasch (1960, pp. 109–125) shows that Maxwell's 1876 analysis of the concepts of mass and force can be expressed mathematically as a multiplicative law and in the form of a model identical with the models Rasch (1960) had previously developed for measuring reading ability. Rasch concludes that:

> *Where this law* [relating reading ability and text difficulty] *can be applied it provides a principle of measurement on a ratio scale of both stimulus parameters and object parameters, the conceptual status of which is comparable to that of measuring mass and force*. Thus . . . the reading accuracy of a child . . . can be measured with the same kind of objectivity as we may tell its weight . . . . (p. 115)

The same kind of objectivity is obtained in the measurement of both the child's weight and his or her reading ability, but not by leaping, in the manner of Galileo, Descartes, Newton and positivism, past the presupposition that either weight or reading ability exists quantitatively.

Instead, Rasch sets up pairs of variables, mass and force, or reading ability and text difficulty, in conjoint relationships through which the variables take their measures each by means of the other. In Rasch's (1960) own words:

> The turning point in this discussion [of the formal properties of the model parameters] was the realization that it did not seem feasible to introduce the two concepts [ability and difficulty] separately, we had to formalize them simultaneously. And in trying to do so we had to introduce a fresh point of view which, however, almost offered itself (ch. V, 7 and 9). . . . It may be illuminating to see that the same principle of introducing two concepts simultaneously, one by means of the other, as it were, and vice versa, without getting into any logical circle, has also proved indispensable to classical physics. (p. 110)

Rasch at this point proceeds into Maxwell's analysis of mass and force, without explicitly intending to make a mathematical demonstration of the function of the hermeneutic circle in the measurement processes of both the natural and human sciences, but accomplishing that demonstration nonetheless.

That is, at the same time that he formulated a mathematical model structurally capable of encompassing both Newton's second law of motion and reading ability, Rasch unknowingly provided a mathematical formalization of the circle Heidegger (1962) held to hide the ‘positive possibility of the most primordial kind of knowing’ (p. 195), the circle through which the scientific theme is made secure by working out the fore-structures of understanding in terms of the things themselves. In the same vein, Heelan’s (1983a; Ihde, 1991, 1998) hermeneutic of instrumentation in the natural sciences builds on Ricoeur’s (1981, p. 210) point that the model of the text provides a sufficient basis for a kind of objectivity that is appropriate to both the human and the natural sciences. The hermeneutic context requires that ‘the first declaration of hermeneutics is to say that the problematic of objectivity presupposes a prior relation of inclusion which encompasses the allegedly autonomous subject and the allegedly adverse object’ (Ricoeur, 1981, p. 105). But when a sign comes to belong together with a particular kind of thing, any spoken or written instance of that sign can be understood as representing any instance of the thing it points to.

This sign–thing coordination facilitates the figure–meaning independence fundamental to academic metaphysics, and is the basis for the objectivity of the model of the text, in a paradigm of reading that dominates both the humanities and the sciences. Ricoeur (1981) explains that:

> This paradigm draws its main features from the status of the text itself as characterized by (1) the fixation of the meaning, (2) its dissociation from the mental intention of the author, (3) the display of non-ostensive references, and (4) the universal range of its addressees. These four traits taken together constitute the ‘objectivity’ of the text. From this ‘objectivity’ derives a possibility of *explaining* which is not derived in any way from another field, that of natural events, but which is congenial to this kind of objectivity. Therefore there is no transfer from one region of reality to another—let us say, from the sphere of facts to the sphere of signs. It is within the same sphere of signs that the process of objectification takes place and gives rise to explanatory procedures. And it is within the same sphere of signs that explanation and comprehension are confronted. (p. 210)

Ricoeur’s sense of the sphere of signs as providing a deeper and truer source for the process of objectification than the sphere of facts is supported by the work of philosophically minded physicists, such as Nobel Prize winners Niels Bohr and John Wheeler.

For instance, Petersen (1968) quotes Bohr regarding ‘the epistemological primacy of the conceptual framework’ in science, saying that

> the attitude to the language–reality problem that it suggests is perhaps best expressed in a remark Bohr once made in a discussion. He was forcefully stressing the primacy of language: ‘Ultimately, we human beings depend on our words. We are hanging in language.’ When it was objected that reality is more fundamental than language and lies beneath language, Bohr

> answered: 'We are suspended in language in such a way that we cannot say what is up and what is down'. (pp. 187–188)

Gadamer's (1989) comment that 'the process of understanding moves entirely in a sphere of a meaning mediated by the verbal tradition' (p. 391) aptly sums up Bohr's sense of how we are suspended in language. Bohr's reasons for saying this stem from what Wheeler calls the 'deepest lesson' of quantum mechanics, Bohr's insight that 'no elementary quantum phenomenon is a phenomenon until it is a recorded (observed) phenomenon' (Wheeler & Zurek, 1983, pp. xvi, 184–185). This lesson arises as the outcome of a search for some kind of unambiguous information on which interpretations of quantum measures may be based, a search that has provisionally concluded with the fixation of meaning in writing (Wheeler & Zurek, 1983).

But in extending the suspension in language to such an explicit reliance on legible text, physics is appealing to hermeneutics insofar as 'everything written is, in fact, the paradigmatic object of hermeneutics' (Gadamer 1989, p. 394). Heelan (1983a, 1983c) takes this principle into the philosophy of science in terms of a hermeneutic of instrumentation involving what he calls 'readable technology', and Ihde (1998) similarly construes the interpretation of visual images in science. Ackermann (1985) also analyzes science in terms of a dialectic of instrumentation, but makes the data text the locus of interpretive sensibility. Wheeler (in Wheeler & Zurek, 1983, p. 185) explicitly accepts what Gadamer (1989, pp. 362–379) calls the 'hermeneutic priority of the question' when he says that, in contemporary physics, the 'answer we get depends on the question we put, the experiment we arrange, the registering device we choose. We are inescapably involved in bringing about that which appears to be happening.' In their study of scales, meaningfulness and quantitative laws, Falmagne and Narens (1983) similarly remark both that 'There is a slow but steady recognition of the odd fact that the language itself which we use in our quantitative description of the world, conditions in a subtle way the image that we obtain', and that 'In quoting quantitative empirical laws, scientists frequently neglect to specify the various scales entering in the equations' (p. 287). In each of these approaches to the problem of objectivity, the hermeneutic-mathematical principle comes to bear: we learn only through symbols and questions that transparently present to us what we already know.

In Ricoeur's terms, Rasch has set up a dialectic of belonging and distanciation between ability and difficulty by requiring and checking to see that test and survey questions and answers serve to effectively mediate mutual understanding by exhibiting the four traits constitutive of the objectivity of the text. In Rasch's terms, test and survey data must simultaneously support both an evidence-based conjoint ordering of questions and answers (or fusion of horizons), and summarization into the statistically sufficient ordinal scores necessary for parameter separation and invariant ratio/interval measures. Thus, in this convergence of independent developments in hermeneutics and mathematical theory, the implicit hermeneutic-mathematical metaphysics of academic values is seen to structure the development of the foundations of simultaneous conjoint measurement theory (Andersen, 1980; Andrich, 1988; Fischer, 1974, 1995; Krantz, et al., 1971; Luce et al., 1990; Luce & Tukey, 1964; Michell, 1990, 1999; Rasch, 1960, 1961, 1977; Suppes et al., 1989; Suppes & Zinnes, 1963; Wright, 1968, 1977, 1985, 1999). The process of reversing Descartes' self-betrayal and filling the void left by Galileo's 'fateful omission' (Husserl, 1970) is therein begun. In the same way that Galileo's 'geometry of idealities was preceded by the practical art of surveying, which knew nothing of idealities' (Husserl, 1970, p. 49), so, too, are abstract human sciences of mathematical idealities being today preceded by surveys and tests that remain tied to particular samples of questions and answers, and so know nothing of idealities.

Rasch's disclosure of the positive possibilities of primordial, mathematical knowing hidden within the hermeneutic circle prompts a more complete recognition of the fact that a positive theory of unified science need not be positivist. Positivism breaks the hermeneutic circle and its dialectic of question and answer, advancing into logic and quantitative science on the basis of unexamined presuppositions about the nature of existence and knowledge. If it were in fact possible to arrive at a positivist theory and practice of measurement, one that effectively results in a positive human science of non-arbitrary reference standard metrics as the mathematical languages researchers think in, why are these not in hand?

Mainstream measurement theory in the human sciences has been nearly completely positivist for most of the last century, in that (1) it left unexamined the criteria it necessarily assumed in recognizing existence and knowledge, and in checking its mathematical presuppositions, and so (2) it proceeded directly into quantification without any metaphysical, metaphorical, dialectical or hermeneutical detours involving simultaneous conjoint circles. If that approach were ever to result in the realization of positive human sciences mathematically akin to the natural sciences in the development and dissemination of quantitative languages scientists actually think in together, would not that goal have been achieved long ago?

To repeat: a *positive theory of unified science need not be positivist*. Rasch and others who have formulated similar 'simultaneous, conjoint' approaches to measurement, such as Luce and Tukey (1964), in particular, have, in contrast with positivism, developed positive science within a context in principle open to deconstructive moments situated appropriately alongside reductive and constructive moments, within the larger phenomenological, or ontological, method (Fisher, 1999b, 2003b; Heidegger, 1982).

The theoretical openness to deconstructive moments is put into practice in software commonly used for estimating measures and instrument calibration values (Andrich, Lyne, Sheridan, & Luo, 2000; Linacre, 1995; Wright & Linacre, 2003; Wu, Adams, & Wilson, 1998; the same openness may also be facilitated by software with which the author is unfamiliar, such as Allerup & Sorber, 1977; ASC, 1996; Fischer, 1997; Glas & Ellis, 1995; Gustafsson, 1979; Kelderman & Steen, 1988; Ludlow, 1992; Schulz, 1988; R.M. Smith, 1991; Verhelst, 1993). These programs' analyses provide not only individualized quantitative numeric indications of amount but also an error estimate and as many as five or six additional evaluations of the extent to which the metaphysical requirement of figure–meaning coordination and independence has been achieved. In accord with Putnam's (1979) sense of mathematical truth as 'ultimate goodness of fit' (p. 394), these evaluations take the form of statistical tests of the model fit (R.M. Smith, 1985, 1996, 2000), construct validity (Fisher, 1994), conjoint additivity (Luce & Tukey, 1964) or sufficiency (Andersen, 1977; R.A. Fisher, 1922) of the observations from which the measures are derived (Fischer, 1995; Wright, 1999).

In this way, the sufficiency of every supposed quantitative reduction is challenged from multiple perspectives, providing opportunities for theoretical reflection on the failures of invariance that may lead to new discoveries (Fisher, 1999b, 2003a, 2003b; Kuhn, 1977; Wimsatt, 1981) or to the correction of clerical errors, the improvement of the instrument or the minimization of bias (Wright, 1977; Wright & Stone, 1979). To appropriate Derrida's language (in Wood & Bernasconi, 1988, pp. 88–89), the error and fit statistics routinely provided for every measure and calibration allow researchers to place themselves at the points at which the thing signified is not easily separable from the signifier. Conjoint measurement thereby integrates qualitative studies of anomalous observations, of what instruments measure, and of quantitative existence, with an explicit theory of the additive relations required for quantitative knowledge.

There can be little doubt that fundamental measurement theory has correctly identified the mathematical criteria necessary for the non-arbitrary coordination of numeric figures with meaningful quantities, in any science, given:

1. the wide variety of proofs and derivations supporting and leading to Rasch's separability theorem and the theorems of fundamental measurement (Andersen, 1977; Fischer, 1995; Krantz et al., 1971; Luce et al., 1990; Luce & Tukey, 1964; Roskam & Jansen, 1984; Suppes et al., 1989; Suppes & Zinnes, 1963; Wright, 1985, 1999);
2. Rasch's (1960, pp. 110–115) exposition of the mathematical similarity of the laws and models feasible in both the natural and human sciences;
3. highly correlated and linear empirical comparisons of Rasch-calibrated interval measures (based in ordinal observations) of weight, length and

distance with the respective grams, centimeters and mapped meters (Choi, 1997; Fisher, 1988, pp. 389–407; Moulton, 1993); and

4. independent experimental results showing the same construct to emerge across different samples from a common population measured with the same instrument, or with different instruments intended to quantify the same variable (Fisher, 1997a, 1997b, 1999a; Fisher et al., 1995, 1997).

Many hundreds, if not thousands, of probabilistic conjoint measurement analyses have been conducted and reported over the last 40 years. One study conducted in the early 1970s involved 2,644 test questions from seven widely used standardized reading tests, over 300,000 students, and 1,650 elementary schools in all 50 US states (Jaeger, 1973, in Engelhard, 2001; also see Rentz & Bashaw, 1977). Extensive study established the interchangeability of the measures from those seven tests, but this did not, however, lead to the widespread availability and acceptance of a universal uniform metric for reading ability. So the question arises as to the sufficiency of these necessary conditions for bringing metaphysically astute measurement into common practice in classrooms, clinics, workplaces, experimental research, and so on (Fisher, 1993, 1995, 1996b). If the rigorous coordination and independence of figure and meaning are mathematically arranged, observed and documented in existing research employing the principles of fundamental measurement theory, why are universal uniform measures virtually unknown in the human sciences?

This question is symptomatic of a fundamental but unresolved issue in the philosophy of science, what Crease (1993) calls ‘the antinomic character of scientific knowledge’ (pp. 164–165). Restating the question so that the antinomy becomes cooperation, Crease suggests a balancing of interests in objective invariant structures with interests in the shaping of those structures by cultural and historical forces. This is, then, the issue of postmodern science manifest at the level of method, where we risk perhaps nonsensical assertions of figure–meaning convergence and separation, and then pay very close attention to the historical shaping and consequences of that invariance and its associated forms of nonsense.

Universal Uniform Measures as the Media of Distributed Cognition

In other words, following Wise (1995), the non-arbitrary, invariant results of multiple Rasch analyses of the same variable, using different instruments on different samples (Fisher, 1997a), the same instrument on different samples (Fisher, 1997b, pp. 360–364; 1999a), or different instruments on the same sample (Fisher et al., 1995, 1997), ought to be functioning as agents of agreement, provoking researchers to recognize and capitalize on the existence of an identified phenomenon that persistently resists tests of its strength and so establishes itself as real. But, despite repeated invitations (e.g. Fisher, 1996b, 1997b, pp. 364–371), researchers have yet to take steps, analogous to those taken in the natural sciences and engineering, toward transforming the accumulating non-arbitrary experimental *agents* of agreement into arbitrary conventional *products* of agreement, that is, reference standard metrics. Universal uniform measures are the realization of the mathematical project in any field that becomes scientific (Heidegger, 1967, p. 93). Accordingly, 'a true union of mathematics and measurement' was not achieved in physics until universal uniform measures were achieved (Roche, 1998, p. 145). Taking up the Socratic metaphor of midwifery, could it be that mathematically viable ideal forms of life are being born in many fields into environments incapable of nurturing them to maturity?

The critique of mainstream measurement practice traces to positivist and Pythagorean roots the assumption that assigning numbers to observations is sufficient for measurement (Fisher, 1992, 1994, 2003a, 2003b; Michell, 1999). Ignoring the hermeneutic-mathematical, technological and social roots of measurement in the natural sciences prevented neither additive, generalized measurement nor metrology from taking hold, though ignoring these roots has undoubtedly prevented the full union of mathematics and measurement from being realized as often as it could be in the natural sciences (Fisher & Markward, 2000; Jacobs, 2001; Markward, 2001; Markward, Fisher, & Keats, 2002; Seccombe, 2001), as well as in the human sciences. It is likely that these issues will need to be addressed directly and decisively before measurement will begin to realize its mathematical potential in many sciences. Testing data for additive relations takes the fundamental hermeneutic-mathematical step of figure–meaning coordination. Designing instruments so as to maximize the likelihood of obtaining additive relations (Fisher, 2000b; Haladyna, 1994; Linacre, 1993, 1997; Woodcock, 1999) takes the fundamental technological step of embodying that coordination in a portable and simplified experimental apparatus, usable by persons lacking full understanding of the device.

The hermeneutic-mathematical and technological steps of research determine the extent to which unit amounts of the variable of interest add up in a way that can be usefully represented by numbers. This determination is the express purpose of fundamental measurement theory in general, and of Rasch's separability theorem in particular. Numbers are the purest form of mathematical symbols only to the extent that their meanings are exhausted by their positions along a uniformly divisible and ordered continuum relative to amounts of the thing measured (Michell, 1990, 1994; Wright, 1985, 1999). In a thoroughly metaphysically astute manner, fundamental measurement theory experimentally evaluates data, testing for figure–meaning convergence and separation en route to calibrating instruments and estimating measures.

Contrary to the expectations associated with what seems to be an unstated but common assumption, these two steps have not led to the wide dissemination of universal uniform measures as the quantitative languages of the human sciences. This may be due to the effect of lingering positivist habits of mind. One of the implications of the Pythagorean sense of the world as number, and of the positivist willingness to leap to the quantitative without thinking through its metaphysical presuppositions, is that the supposed self-evident givenness of quantity in nature is sufficient in and of itself to integrate the implied numerical relations into research applications and/or everyday life. A significant literature in the social studies of science suggests that this is not the case, and that there are specific processes through which mathematical uniformities become distributed throughout society and are made traceable to reference standards (Callon, 1995; Knorr Cetina, 1995; Latour, 1987, 1993, 1994, 1995, 1999; Mendelsohn, 1992; O'Connell, 1993; Wise, 1995).

There thus appears to be as much of a need in the human sciences as in the natural sciences for agreed-upon procedural standards regarding the intra- and inter-laboratory tests necessary for the dissemination of universal uniform measures (Fisher, 1996a). Following Wise (1995; Widmalm, 1995), real phenomena compel recognition of their properties and so function as agents of agreement when independent studies conducted by different researchers in different laboratories using different instruments on different samples converge on very similar, though as yet not formally coordinated, results. Intra-laboratory ruggedness tests are well recognized as the first phase of metrological study (Wernimont, 1977, 1978).

In other words, mathematical variables are socially constructed in one way to the extent that independent researchers' investigations become phenomenologically coordinated by what Gadamer (1989), connecting with Hegel, calls 'the activity of the thing itself' (pp. 460, 464). When thinking unfolds 'what consistently follows from the subject matter itself' (Gadamer, 1989, p. 464), as it steadily resists tests of its strength (Ihde, 1991; Latour, 1987, 1994; Wimsatt, 1981), and a preliminary convergence of thing and thought, of signified and signifier, is achieved in metrology's first phase, the non-arbitrary phenomenological uniformity of the variable offers itself for routinization via inter-laboratory trials, metrology's second phase (Mandel, 1977, 1978). Mathematical variables are socially constructed in a second way, then, when the audience gathered around a consistently produced phenomenon in turn seeks to enlarge the audience, enrolling other researchers, practitioners, regulators, manufacturers, and so on, by translating their interests so that they coincide and fuse with the interests that give rise to the thing itself.

Quantification is provisionally complete only when the arbitrary conventions of numerical unit and range, tolerable error and data quality are expressed as products of the participating scientists' agreement, enhancing the clarity of their communications, the interpretability of their research and the uniformity of their commercial products and transactions. As Latour (1987) points out:

> The predictable character of technoscience is entirely dependent on its ability to spread networks further. . . . Facts and machines are like trains, electricity, packages of computer bytes or frozen vegetables: they can go everywhere as long as the track along which they travel is not interrupted in the slightest. This dependence and fragility is not felt by the observer of science because 'universality' offers them the possibility of applying laws of physics, of biology, or of mathematics everywhere *in principle*. It is quite different *in practice*. (pp. 249–250)

So, though it is possible in principle to telephone anyone anywhere, it is not possible to phone someone lacking a phone or someone who has a phone but no connection to the network; Ohm's law may be constant across the entire universe, but it cannot be demonstrated without a voltmeter, wattmeter, ammeter and power source; a machine should work the same way any time and any place for any operator, but often cannot be fixed without the right tools; and so on. In other words, the task of research into a particular variable is not fully mathematical until the tools of that research are formally coordinated into a structured sign system capable of functioning as the medium for collective, distributed cognition.

The metaphorical, numerical, geometrical and dramatic figures, written texts and other readable technologies associated with the production of reproducible, recognizable effects, which bring them into view and give them life as abstractions independent from the particular representations involved, must be stable, mobile and transferable, and combinable (Latour, 1987, p. 223). Cartography is the example Latour chooses to show the manner in which representations of new territory can be effectively fixed as intellectual capital in a transportable form that enables others who have never seen the territory nonetheless to be prepared with the information they need to re-cognize it when they see it. Maps drawn by one group of people employing the recognizable symbols, figures and proportions of geometry can be read by another entirely different group of people trained in interpreting the language of the map so that they see virtually what the first group saw actually. Accordingly, recent studies in the history and philosophy of science and literacy research have come to focus on the ways in which various forms of intellectual capital are combined or translated across linguistic markets (Galison, 1999; Luke, 2003).

Metrology is the discipline that maintains the quality of the intellectual capital over all the translations, and through which scientific results are made universal by means of generalizations from particulars (Latour, 1987; O'Connell, 1993; Wise, 1995). Metrology routinizes the means of producing in the laboratory world what are perhaps rare phenomena in the outside world, and then it exports the routinized phenomena of the laboratory world in a prepackaged technical microcosm, which, to continue to function in the laboratory's prescribed manner, must remain connected with the laboratory. For instance:

> There is a continuous trail of readings, checklists, paper forms, telephone lines, that tie all the clocks together. As soon as you leave this trail, you start to be *uncertain* about what time it is, and the only way to regain certainty is to get in touch again with the metrological chains. Physicists use the nice word *constant* to designate these elementary parameters necessary for the simplest equation to be written in the laboratories. These constants, however, are so inconstant that the US, according to the National Bureau of Standards, spends 6 per cent of its Gross National Product, that is, three times [actually, in the early 1990s, about 2.3 times (Subcommittee on Research, 1996)] what is spent on R & D, just to maintain them stable! . . . That much more effort has to be invested in extending science than in doing it may surprise those who think it is naturally universal. (Latour, 1987, pp. 251, 252)

The effort and expense involved in creating and maintaining standards are paid back in a number of ways. Standards are valuable for enhancing safety, as became painfully evident in Baltimore in 1904 when a large fire attracted emergency units from several different cities and it was discovered that few of the hose couplings matched the threads in the hydrants (Cochrane, 1966, pp. 84–6), turning a relatively minor disaster into a catastrophe. Standards are valuable economically, for many reasons, including safety, the interchangeability of parts, the regional expansion of markets and the high returns on investment produced by metrological quality improvement studies. One set of twelve such studies (National Institute for Standards and Technology, 1996) averaged returns of over 140 percent. Standards probably also play a role in resolving theoretical disputes in science by enabling conclusive determinations of the reproducibility of experimental results (Fisher, 2003a).

Measurement standards could similarly play important roles in enhancing the safety, economic value and theoretical progress made in applications in the human sciences. For instance, the recent Institute of Medicine report *To Err is Human* (Kohn, Corrigan, & Donaldson, 2000) suggests that a large proportion of errors in health care could be avoided by including more information from the patient in various decisions. Standardized measures of health status, quality of care, disease severity, and so on, could provide essential checks on the extent to which the provider and the patient agree on the problem and on the effectiveness of the treatment.

The economic and scientific impact of universal uniform measures of health status may become most tangible in quality improvement projects and clinical trials where the comparative effectiveness of different treatment modalities may be determined more efficiently and generally than can be accomplished currently using non-linear, ordinal, sample-dependent scores.

But for the shift from the current paradigm to the new one to occur, a broad, comprehensive appreciation will be required of the vital roles played by demanding mathematical models and by the fixation of meaning in uniformly transferable forms. To be convincing and demonstrably effective, the metaphysical principles of first philosophy must be shown to be capable of resolving at least two outstanding problems, including the resolution of the subject–object dichotomy that structures the Cartesian definition of modern science, and the ambiguity of metaphor. If viable solutions to these problems might be suggested by following through on a path of thinking that pursues Western mathematical metaphysics to its practical and logical consequences, as is attempted in work currently in process, then perhaps a plan for an effective and comprehensive transformation of the human sciences into caring sciences might come into view.

Concluding Comments

Mathematical languages that researchers, practitioners, managers, consumers, students and so on, think in together, facilitated via distributed networks of instruments traceable to reference standards, and facilitating the emergence of a ‘new metrological culture’ (Roche, 1998, p. 148), come into being only to the extent that the ‘full union of mathematics and measurement’ (Roche, 1998, p. 145) is realized. For mathematics and measurement to be integrated fully, the quantities of experimental research must, first, act as agents of agreement among investigators by representing additive amounts of the variables of interest that compel recognition of them as objectively existing, as determined via multiple independent experimental studies evaluating the additive structure of divisible magnitudes. Second, these quantities must in turn be shaped as the conventional products of investigators’ agreement, so that different brands or configurations of instruments measuring the same thing do so in the same universal uniform metric.

In this context of the full union of mathematics and measurement, Kuhn (1977) ‘ventured the following paradox’: ‘The full and intimate quantification of any science is a consummation devoutly to be wished. Nevertheless, it is not a consummation that can effectively be sought by measuring’ (p. 221). The paradox is now less paradoxical, since we see that coordinated social activities, and not the activity of measuring itself, are what consummate the full and intimate quantification of a science. Kuhn (1977) further suggests that ‘maturity comes most surely to those who know how to wait’ (p. 221), implying that, in the manner of the emergence of professional metrology in the 19th century, full quantification may have to happen by itself in its own time.

Some may wait, but others may act (Fisher, 1993, 1995, 1996b, 2000a). Michell shows that the vast majority of measures made in the human sciences do not live up to the real meaning of the word ‘measurement’, and so are non-linear, ordinal, scale- and sample-dependent numbers that provide little or no quantitative expression to their intended variables. It is by now clear, as suggested by Kuhn in the passage just quoted, that measurement of this sort will not lead to full and intimate quantification in the sense of the compelling agents and products of agreement available in the natural sciences. But many researchers employing rigorous measurement methods have already been working for decades in uncoordinated isolation from one another, publishing results that provide largely unrecognized evidence of construct congruence across instruments, respondent samples, time and space (Fisher, 1997a, 1997b, 1999a).

As Kuhn suggests, maturity may come to these researchers if they wait, since, as they continue working in isolation, the hermeneutic-mathematical uniformities of the things themselves will eventually become so apparent that their common structures will impose themselves upon the social organization of the research. That is, in the same way that ‘the mathematical project of Newtonian bodies leads to the development of a certain “mathematics” in the narrow sense’ of universal uniform numerical measurement (Heidegger, 1967, p. 93), so, too, will the rigorously mathematical projects of human abilities, attitudes and performances measured via Rasch’s models eventually set the stage for the development of reference standard metrics in the psychosocial sciences (Fisher, 1995, 1996b, 2001).

But maturity may not be reserved solely for those who know how to wait. On the contrary, innovators will recognize opportunities for social, political and/or economic gain in the fact that different ways of questioning a variable can sometimes (and sometimes not) be coordinated into a single, unified mathematical sign system, and they will act accordingly. In contrast with modernity’s hidden metaphysics, incomplete hermeneutics and mathematics, and haphazard quantifications, postmodern science lays out the entire continuum of the hermeneutic-mathematical enterprise. The postmodern perspective is cognizant of the full spectrum of research, from the experimental test of quantitative status through to the social enrollment of end users and the construction of the legal, information and financial infrastructure through which they will be linked into regional, national and global economies. Where each introduction of a new instrument in the current methodological context adds ever more levels of incommensurable sample- and scale-dependent score units to a growing Tower of Babel, metaphysically astute and metrologically informed measurement in postmodern science will determine the extent to which common quantitative languages can be conceived, born and nurtured for each form of living capital. Universal uniform measurement thus requires an explicit articulation of the structure of quantitative regularity in a model, and comparative checks of observed data against modeled expectations. These issues will be taken up in forthcoming works.

Note

1. For a chart of cumulative annual counts of Rasch papers indexed on MEDLINE, see Fisher (2002a).

References

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

Adams, R.J., & Wright, B.D. (1994). When does misfit make a difference? In M. Wilson (Ed.), *Objective measurement: Theory into practice* (Vol. 2, pp. 244–270). Norwood, NJ: Ablex.

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

Andersen, E.B. (1973). A goodness of fit test for the Rasch model. *Psychometrika*, *38*, 123–140.

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.

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

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

Andrich, D.A. (1989a). Constructing fundamental measurements in social psychology. In J.A. Keats, R. Taft, R.A. Heath, & S.H. Lovibond (Eds.), *Mathematical and theoretical systems: Proceedings of the 24th International Congress of Psychology of the International Union of Psychological Science* (Vol. 4, pp. 17–26). Amsterdam: North-Holland.

Andrich, D.A. (1989b). Distinctions between assumptions and requirements in measurement in the social sciences. In J.A. Keats, R. Taft, R.A. Heath, & S.H. Lovibond (Eds.), *Mathematical and theoretical systems: Proceedings of the 24th International Congress of Psychology of the International Union of Psychological Science* (Vol. 4, pp. 7–16). Amsterdam: North-Holland.

Andrich, D.A. (2002). Understanding resistance to the data–model relationship in Rasch's paradigm: A reflection for the next generation. *Journal of Applied Measurement*, *3*(3), 325–59.

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

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

Bayley, S. (2000). Measuring customer satisfaction: Comparing traditional and latent trait approaches using the Auditor General's survey. *Evaluation Journal of Australasia*, *1*(1), 8–17.

Bazerman, C. (1994). *Shaping written knowledge: The genre and activity of the experimental article in science*. Madison, WI: University of Wisconsin Press.

Bergstrom, B.A., & Lunz, M.E. (1999). CAT for certification and licensure. In F. Drasgow & J.B. Olson-Buchanan (Eds.), *Innovations in computerized assessment* (pp. 67–91). Mahwah, NJ: Erlbaum.

Bond, T., & Fox, C. (2001). *Fundamental measurement in the human sciences: Applying the Rasch model.* Hillsdale, NJ: Erlbaum.

Bourbaki, N. (1971). The architecture of mathematics. In F. Le Lionnais (Ed.), *Great currents in mathematical thought* (pp. 23–36). New York: Dover.

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

Callon, M. (1995). Four models for the dynamics of science. In S. Jasanoff, G.E. Markle, J.C. Petersen, & T. Pinch (Eds.), *Handbook of science and technology studies* (pp. 29–63). Thousand Oaks, CA: Sage.

Campbell, N.R. (1920). *Physics, the elements*. Cambridge: Cambridge University Press.

Campbell, S.K., Wright, B.D., & Linacre, J.M. (2002, Summer). Development of a functional movement scale for infants. *Journal of Applied Measurement*, *3*(2), 190–204.

Carlson, S., Andrews, M., & Bickel, G. (1999, Feb). Measuring food insecurity and hunger in the United States: Development of a national benchmark measure and prevalence estimates. *Journal of Nutrition*, *129*(2S Suppl), 510s–516s.

Carnap, R. (1959). The elimination of metaphysics through logical analysis of language. In A.J. Ayer (Ed.), *Logical positivism* (pp. 60–81). New York: Free Press. [Reprinted in Murray, M. (1978), *Heidegger and modern philosophy* (pp. 23–34). New Haven, CT: Yale University Press.]

Choi, S.E. (1997). Rasch invents 'Ounces.' *Rasch Measurement Transactions*, *11*(2), 557 [http://www.rasch.org/rmt/rmt112.htm£Ounces].

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 & L.J. van der Kamp (Eds.), *Advances in psychological and educational measurement* (pp. 233–245). New York: Wiley.

Cliff, N. (1992). Abstract measurement theory and the revolution that never happened. *Psychological Science*, *3*, 186–190.

Cliff, N. (1993). What is and isn't measurement. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 53–90). Hillsdale, NJ: Erlbaum.

Cochrane, R.E. (1966). *Measures for progress: A history of the National Bureau of Standards*. Washington, DC: US Dept. of Commerce.

Connolly, A.J., Nachtman, W., & Pritchett, E.M. (1971). *Keymath: Diagnostic Arithmetic Test*. Circle Pines, MN: American Guidance Service.

Coye, M.J. (2001, November/December). No Toyotas in health care: Why medical care has not evolved to meet patients' needs. *Health Affairs*, *20*(6), 44–56.

Crease, R. (1993). *The play of nature*. Bloomington: Indiana University Press.

Crosby, A.W. (1997). *The measure of reality: Quantification and Western society, 1250–1600*. Cambridge: Cambridge University Press.
Danziger, K. (1985). The methodological imperative in psychology. *Philosophy of the Social Sciences*, *15*(1), 1–13.
Dawson, T.L. (2002, Summer). A comparison of three developmental stage scoring systems. *Journal of Applied Measurement*, *3*(2), 146–189.
Derrida, J. (1978). Structure, sign and play in the discourse of the human sciences. In *Writing and difference* (A. Bass, Trans.; pp. 278–293). Chicago, IL: University of Chicago Press.
De Soto, H. (2000). *The mystery of capital: Why capitalism triumphs in the West and fails everywhere else*. New York: Basic Books.
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. (2001). Historical view of the influences of measurement and reading theories on the assessment of reading. *Journal of Applied Measurement*, *2*(1), 1–26.
Falmagne, J.-C., & Narens, L. (1983). Scales and meaningfulness of quantitative laws. *Synthese*, *55*, 287–325.
Fischer, G.H. (1974). *Einführung in die theorie psychologischer tests* [Introduction to mental test theory]. Berne: Huber.
Fischer, G.H. (1995). Derivations of the Rasch model. In G. Fischer & I. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 15–38). New York: Springer-Verlag.
Fischer, G.H. (1997). Structural Rasch models: Some theory, applications, and software. In M. Wilson, G. Engelhard, Jr., & K. Draney (Eds.), *Objective measurement: Theory into practice* (Vol. 4, pp. 185–208). Norwood, NJ: Ablex.
Fischer, G.H., & Molenaar, I. (Eds.). (1995). *Rasch models: Foundations, recent developments, and applications*. New York: Springer-Verlag.
Fisher, A.G. (1993, April). The assessment of IADL motor skills: An application of many-faceted Rasch analysis. *American Journal of Occupational Therapy*, *47*(4), 319–329.
Fisher, R.A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London*, *A*, *222*, 309–368.
Fisher, W.P., Jr. (1988). Truth, method, and measurement: The hermeneutic of instrumentation and the Rasch model (Doctoral dissertation, University of Chicago). *Dissertation Abstracts International*, *49*, 0778A.
Fisher, W.P., Jr. (1992). Objectivity in measurement: A philosophical history of Rasch's separability theorem. In M. Wilson (Ed.), *Objective measurement: Theory into practice* (Vol. 1, pp. 29–58). Norwood, NJ: Ablex.
Fisher, W.P., Jr. (1993). Scale-free measurement revisited. *Rasch Measurement Transactions*, *7*(1), 272–3 [http://www.rasch.org/rmt/rmt71.htm].
Fisher, W.P., Jr. (1994, August 20–4). The Rasch debate: Validity and revolution in educational measurement. In M. Wilson (Ed.), *Objective measurement: Theory into practice* (Vol. 2, pp. 36–72). Norwood, NJ: Ablex.

Fisher, W.P., Jr. (1995). Opportunism, a first step to inevitability? *Rasch Measurement Transactions*, *9*(2), 426 [http://www.rasch.org/rmt/rmt92.htm].

Fisher, W.P., Jr. (1996a, October). *Measurement standards relevant to ASTM 1384 on the content and structure of the electronic health record.* Unpublished paper presented at the ASTM E31 Committee on the Electronic Health Record, Washington, DC.

Fisher, W.P., Jr. (1996b, Winter). The Rasch alternative. *Rasch Measurement Transactions, 9*(4), 466–467 [http://www.rasch.org/rmt/rmt94.htm].

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). What scale-free measurement means to health outcomes research. *Physical Medicine & Rehabilitation State of the Art Reviews, 11*(2), 357–373.

Fisher, W.P., Jr. (1999a). 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. (1999b). *Reducible or irreducible? Sufficiency and invariance in psychological measurement.* Paper presented at the annual meeting of the American Psychological Association, Boston.

Fisher, W.P., Jr. (2000a). Objectivity in psychosocial measurement: What, why, how. *Journal of Outcome Measurement*, *4*(2), 527–563.

Fisher, W.P., Jr. (2000b). Survey design recommendations. *Popular Measurement*, *3*(1), 58–9.

Fisher, W.P., Jr. (2001). Review of John Roche's *The mathematics of measurement: A critical history*. *Journal of Applied Measurement*, *2*(4), 426–440.

Fisher, W.P., Jr. (2002a, Winter). Growth of Rasch publications in the medical literature. *Rasch Measurement Transactions*, *16*(3), 884 [http://www.rasch.org/rmt/rmt163.zip].

Fisher, W.P., Jr. (2002b, Spring). 'The mystery of capital' and the human sciences. *Rasch Measurement Transactions*, *15*(4), 854 [http://www.rasch.org/rmt/rmt154j.htm].

Fisher, W.P., Jr. (2003a). Consequences of mathematical metaphysics for metrology in the human sciences. In A. Morales (Ed.), *Renascent pragmatism: Studies in law and social science* (pp. 118–153). Brookfield, VT: Ashgate.

Fisher, W.P., Jr. (2003b). Mathematics, measurement, metaphor and metaphysics I. Implications for method in postmodern science. *Theory & Psychology*, *13*, 753–790.

Fisher, W.P., Jr., Eubanks, R.L., & 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., Taylor, P., Kilgore, K.M., & Kelly, C.K. (1995). Rehabits: A common language of functional assessment. *Archives of Physical Medicine and Rehabilitation*, *76*, 113–122.

Fisher, W.P., Jr., & Markward, N. (2000). Applying fundamental measurement to problems in cancer genetics: Mutations as signature violations of sufficiency and model fit [abstract]. *Annals of the Academy of Studenica*, *3*, 23.

Fisher, W.P., Jr., & Wright, B.D. (Eds.). (1994). Applications of probabilistic conjoint measurement [Special Issue]. *International Journal of Educational Research, 21*(6), 557–664.

Gadamer, H.-G. (1989). *Truth and method* (2nd rev. ed.; G. Barden & J. Cumming, Eds.; J. Weinsheimer & D. Marshall, rev. Trans.). New York: Crossroad.

Galison, P. (1999). Trading zone: Coordinating action and belief. In M. Biagioli (Ed.), *The science studies reader* (pp. 137–160). New York: Routledge.

Gehlert, S., & Chang, C. (1998). Factor structure and dimensionality of the multidimensional health locus of control scales in measuring adults with epilepsy. *Journal of Outcome Measurement*, *2*(3), 173–190.

Glas, C.A.W., & Ellis, J.L. (1995). *RSP 1.0: Rasch scaling program.* Arnhem: CITO.

Glas, C.A.W., & Verhelst, N.D. (1995). Tests of fit for polytomous Rasch models. In G. Fischer & I. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 325–352). New York: Springer-Verlag.

Glazebrook, T. (2000). *Heidegger's philosophy of science*. New York: Fordham University Press.

Granger, C.V., & Gresham, G.E. (Eds). (1993, August). New developments in functional assessment [Special Issue]. *Physical Medicine and Rehabilitation Clinics of North America*, *4*(3), 417–611.

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

Guttman, L. (1985). The illogic of statistical inference for cumulative science. *Applied Stochastic Models and Data Analysis*, *1*, 3–10.

Haladyna, T.M. (1994, June). A research agenda for licensing and certification testing validation studies. *Evaluation & the Health Professions*, *17*(2), 242–256.

Heelan, P. (1983a, June). Natural science as a hermeneutic of instrumentation. *Philosophy of Science*, *50*, 181–204.

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

Heelan, P. (1983c). *Space-perception and the philosophy of science*. Berkeley: University of California Press.

Heidegger, M. (1962). *Being and time* (J. Macquarrie & E. Robinson, Trans.). New York: Harper & Row.

Heidegger, M. (1967). *What is a thing*? (W.B. Barton, Jr. & V. Deutsch, Trans.; analytic afterword by E. Gendlin). South Bend, IN: Regnery.

Heidegger, M. (1982). *The basic problems of phenomenology* (A. Hofstadter, Trans.). Bloomington: Indiana University Press.

Heilbron, J.L. (1993). *Weighing imponderables and other quantitative science around 1800.* Historical studies in the physical and biological sciences, Vol. 24 (Supplement), Part I (pp. 1–337). Berkeley: University of California Press.

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

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.

Ihde, D. (1990). *Technology and the lifeworld: From garden to earth*. Bloomington: Indiana University Press.

Ihde, D. (1991). *Instrumental realism: The interface between philosophy of science and philosophy of technology*. Bloomington: Indiana University Press.

Ihde, D. (1998). *Expanding hermeneutics: Visualism in science*. Evanston, IL: Northwestern University Press.

Jacobs, E. (2001, July 31). Point-of-care testing connectivity update—Connectivity Industry Consortium. In E. Jacobs (Chair), *Session 3301: Outcomes & utility of point-of-care testing*. Annual Meeting of the American Association of Clinical Chemists, Chicago, IL.

Jaeger, R.M. (1973). The national test equating study in reading (The Anchor Test Study). *Measurement in Education*, *4*, 1–8.

Kalinowski, A. (1985). Measuring clinical pain. *Journal of Psychopathology and Behavioral Assessment*, *7*, 329–349.

Karabatsos, G. (1998). Analyzing nonadditive conjoint structures: Compounding events by Rasch model probabilities. *Journal of Outcome Measurement*, *2*(3), 191–221.

Karabatsos, G. (2001). The Rasch model, additive conjoint measurement, and new models of probabilistic measurement theory. *Journal of Applied Measurement*, *2*(4), 389–423.

Keats, J.A. (1967). Test theory. *Annual Review of Psychology*, *18*, 217–238.

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

Kelley, P.R., & Schumacher, C.F. (1984, December). The Rasch model: Its use by the National Board of Medical Examiners. *Evaluation & the Health Professions*, *7*(4), 443–454.

Kindig, D.A. (1997). *Purchasing population health*. Ann Arbor: University of Michigan Press.

Kisiel, T. (1973). The mathematical and the hermeneutical: On Heidegger's notion of the apriori. In E.G. Ballard & C.E. Scott (Eds.), *Martin Heidegger: In Europe and America* (pp. 109–120). The Hague: Martinus Nijhoff.

Klauer, K.C. (1995). The assessment of person fit. In G.H. Fischer & I.W. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 97–110). New York: Springer-Verlag.

Klein, H.A. (1974). *The world of measurements: Masterpieces, mysteries and muddles of metrology*. New York: Simon & Schuster.

Kline, P., Laming, D., Lovie, A.D., Luce, R.D., & Morgan, M. (1997). Commentary on Michell, quantitative science and the definition of measurement in psychology. *British Journal of Psychology*, *88*, 385–400.

Knorr Cetina, K. (1995). Laboratory studies: The cultural approach to the study of science. In S. Jasanoff, G.E. Markle, J.C. Petersen & T. Pinch (Eds.), *Handbook of science and technology studies* (pp. 140–66). Thousand Oaks, CA: Sage.

Kohn, L., Corrigan, J., & Donaldson, M. (Eds.). (2000). *To err is human: Building a safer health system*. Washington, DC: National Academy Press.

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

Kuhn, T.S. (1977). The function of measurement in modern physical science. In *The essential tension: Selected studies in scientific tradition and change* (pp. 178–224). Chicago, IL: University of Chicago Press. [Reprinted from Isis, 52(168), 1961, 161–193.]

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

Latour, B. (1993). *We have never been modern.* Cambridge, MA: Harvard University Press.

Latour, B. (1994, May). Pragmatogonies: A mythical account of how humans and nonhumans swap properties. *American Behavioral Scientist*, *37*(6), 791–808.

Latour, B. (1995). *Cogito ergo sumus!* Or psychology swept inside out by the fresh air of the upper deck [Review of Hutchins' *Cognition in the Wild*, MIT Press, 1995]. *Mind, Culture, and Activity: An International Journal*, *3*(1), 54–63.

Latour, B. (1999). *Pandora's hope: Essays on the reality of science studies.* Cambridge, MA: Harvard University Press.

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

Linacre, J.M. (1993). Rasch generalizability theory. *Rasch Measurement Transactions*, *7*(1), 283–284 [http://www.rasch.org/rmt/rmt71.htm].

Linacre, J.M. (1995). *Facets Rasch analysis computer program.* Chicago, IL: MESA Press [www.rasch.org].

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

Linacre, J.M. (1997). Instantaneous measurement and diagnosis. *Physical Medicine and Rehabilitation State of the Art Reviews*, *11*(2), 315–324.

Linacre, J.M. (1998). Detecting multidimensionality: Which residual data-type works best? *Journal of Outcome Measurement*, *2*(3), 266–83.

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

Lobato, J. (2003, January/February). How design experiments can inform a rethinking of transfer and vice versa. *Educational Researcher*, *32*(1), 17–20.

Lobato, J., & Siebert, D. (2002). Quantitative reasoning in a reconceived view of transfer. *Journal of Mathematical Thinking*, *21*(1), 87–116.

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

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

Ludlow, L.H. (1992). *SCALE: A Rasch program for rating scale data.* Chestnut Hill, MA: Educational Research, Measurement, and Evaluation Program, Boston College.

Luke, A. (2003, January–March). Literacy and the Other: A sociological approach to literacy research and policy in multilingual societies. *Reading Research Quarterly*, *38*(1), 132–41.

Lunz, M.E., Bergstrom, B.A., & Gershon, R.C. (1994). Computer adaptive testing. *International Journal of Educational Research*, *21*(6), 623–634.

Lunz, M.E., & Linacre, J.M. (1998). Measurement designs using multifacet Rasch modeling. In G.A. Marcoulides (Ed.), *Modern methods for business research: Methodology for business and management* (pp. 47–77). Mahwah, NJ: Erlbaum.

Lunz, M.E., Wright, B.D., & Linacre, J.M. (1990). Measuring the impact of judge severity on examination scores. *Applied Measurement in Education*, *3/4*, 331–345.

Maasen, S., & Weingart, P. (2001). *Metaphors and the dynamics of knowledge*. London: Routledge.

Mandel, J. (1977, March). The analysis of interlaboratory test data. *ASTM Standardization News*, *5*, 17–20, 56.

Mandel, J. (1978, December). Interlaboratory testing. *ASTM Standardization News, 6*, 11–12.

Maor, E. (1994). *e: The story of a number.* Princeton, NJ: Princeton University Press.

Markward, N. (2001). *A framework for the objective measurement of genomic information.* Unpublished doctoral dissertation, Louisiana State University School of Medicine, Department of Genetics.

Markward, N., Fisher, W.P., Jr., & Keats, B. (2002). *'Family-free' genetic linkage and linkage disequilibrium analyses: A new perspective of the evolutionary dynamics of Friedreich ataxia.* Unpublished ms. New Orleans, LA: LSU Health Sciences Center, Department of Genetics.

Masters, G.N. (1984). Constructing an item bank using partial credit scoring. *Journal of Educational Measurement*, *21*(1), 19–32.

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

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

Mathieu, W.A. (1997). *Harmonic experience: Tonal harmony from its natural origins to it modern expression.* Rochester, VT: Inner Traditions International.

McNamara, T. (1996). *Measuring second language performance.* White Plains, NY: Addison Wesley Longman.

Mendelsohn, E. (1992). The social locus of scientific instruments. In R. Bud & S.E. Cozzens (Eds.), *Invisible connections: Instruments, institutions, and science* (pp. 5–22). Bellingham, WA: SPIE Optical Engineering Press.

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

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

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. (1999). *Measurement in psychology: A critical history of a methodological concept.* Cambridge: Cambridge University Press.

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

Moulton, M. (1993). Probabilistic mapping. *Rasch Measurement Transactions*, *7*(1), 268 [http://www.rasch.org/rmt/rmt71.htm].

National Institute for Standards and Technology. (1996). Economic impacts of research in metrology. In Subcommittee on Research (Ed.), *Assessing fundamental science: A report from the Subcommittee on Research, Committee on Fundamental*

Science. Washington, DC: National Standards and Technology Council [http://www.nsf.gov/sbe/srs/ostp/assess/nstcafsk.htm£Topic%207].

O'Connell, J. (1993). Metrology: The creation of universality by the circulation of particulars. *Social Studies of Science*, *23*, 129–173.

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

Petersen, A. (1968). *Quantum physics and the philosophical tradition.* Cambridge, MA: MIT Press.

Platt, J. (1964, October 16). Strong inference. *Science*, *146*(3642), 347–53.

Putnam, H. (1979). Philosophy of mathematics: A report. In P.D. Asquith & H. Kyburg (Eds.), *Current research in philosophy of science. Proceedings of the PSA critical research problems conference* (pp. 386–398). East Lansing, MI: Philosophy of Science Association.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danmarks Paedogogiske Institut. [Reprinted, 1980, with Foreword and Afterword by B.D. Wright, Chicago, IL: University of Chicago Press.]

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In J. Neyman, (Ed.), *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (pp. 321–333). Berkeley: University of California Press.

Rasch, G. (1966). An item analysis which takes individual differences into account. *British Journal of Mathematical and Statistical Psychology*, *19*, 49–57.

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.

Reckase, M.D. (1989). Adaptive testing: The evolution of a good idea. *Educational Measurement: Issues and Practice*, *8*, 3.

Rentz, R.R., & Bashaw, W.L. (1977, Summer). The National Reference Scale for Reading: An application of the Rasch model. *Journal of Educational Measurement*, *14*(2), 161–179.

Ricoeur, P. (1974). *Political and social essays* (D. Stewart & J. Bien, Eds.). Athens: Ohio University Press.

Ricoeur, P. (1978). The task of hermeneutics (D. Pellauer, Trans.). In M. Murray (Ed.), *Heidegger and modern philosophy* (pp. 141–160). New Haven, CT: Yale University Press. [Reprinted from Ricoeur, P. (1973). The task of hermeneutics. *Philosophy Today*, *17*, 112–28; also in Ricoeur, P. (1981), *Hermeneutics and the human sciences: Essays on language, action and interpretation* (J. B. Thompson, Ed. & Trans.; pp. 43–62). Cambridge: Cambridge University Press.]

Ricoeur, P. (1979). *Main trends in philosophy*. New York: Holmes & Meier. [Reprinted from *Main trends of research in the social and human sciences – Part II*. UNESCO, 1978.]

Ricoeur, P. (1981). *Hermeneutics and the human sciences: Essays on language, action and interpretation*. Cambridge: Cambridge University Press.

Roche, J. (1998). *The mathematics of measurement: A critical history.* London: Athlone.

Roskam, E.E., & Jansen, P.G.W. (1984). A new derivation of the Rasch model. In E. DeGreef & J. van Buggenhaut (Eds.), *Trends in mathematical psychology* (pp. 293–307). Amsterdam: Elsevier.

Rost, J., & von Davier, M. (1994, June). A conditional item-fit index for Rasch models. *Applied Psychological Measurement*, *18*(2), 171–182.

Schulz, E.M. (1988, Winter). Computer program: MFORMS for one-step item banking. *Rasch Measurement Transactions*, *1*(2), 10–11 [http://www.rasch.org/rmt12.htm].

Seccombe, D.W. (2001, July 31). Standardization of laboratory testing: A model for the future? In G. Cembrowski (Chair), *Session 3304: Reshaping proficiency testing: Lessons for laboratorians, providers & regulators*. Annual Meeting of the American Association of Clinical Chemists, Chicago, IL.

Shapin, S. (1994). *A social history of truth: Civility and science in seventeenth-century England.* Chicago, IL: University of Chicago Press.

Smith, E.V., Jr. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement*, *3*(2), 205–231.

Smith, R.M. (1985). A comparison of Rasch person analysis and robust estimators. *Educational & Psychological Measurement*, *45*(3), 433–444.

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

Smith, R.M. (1996). A comparison of methods for determining dimensionality in Rasch measurement. *Structural Equation Modeling*, *3*(1), 25–40.

Smith, R.M. (Ed.). (1997, June). *Physical Medicine & Rehabilitation: State of the Art Reviews: Outcome Measurement*, *11*(2), 261–428.

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

Subcommittee on Research. (Ed.). (1996, July 9). *Assessing fundamental science: A report from the Subcommittee on Research, Committee on Fundamental Science*. Washington, DC: National Science and Technology Council [http://www.nsf.gov/sbe/srs/ostp/assess/].

Sullivan, A.T. (1985). *The seventh dragon: The riddle of equal temperament.* Lake Oswego, OR: Metamorphous.

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

Suppes, P., & Zinnes, J.L. (1963). Basic measurement theory. In R.D. Luce, R.R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (pp. 1–76). New York: Wiley.

Tenenbaum, G. (1999, January–March). The implementation of Thurstone's and Guttman's measurement ideas in Rasch analysis. *International Journal of Sport Psychology, 30*(1), 3–16.

Thorndike, E.L. (1904). *An introduction to the theory of mental and social measurements.* New York: Teacher's College Press.

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

Thurstone, L.L. (1959a). Attitudes can be measured. In *The Measurement of Values* (pp. 215–233). Chicago, IL: University of Chicago Press. [Reprinted from *American Journal of Sociology*, *XXXIII*, 1928, 529–544.]

Thurstone, L.L. (1959b). Psychology as a quantitative rational science. In *The* *measurement of values* (pp. 3–11 Chicago, IL: University of Chicago Press. [Reprinted from *Science*, *LXXXV*, March 5, 1937, 228–232.]

Verhelst, N. D. (1993). *One Parameter Logistic Model computer program.* Arnhem: Cito.

Wegner, D. (1991). Transactive memory in close relationships. *Journal of Personality and Social Psychology*, *61*(6), 923–9.

Weiss, D.J. (1983). *New horizons in testing: Latent trait test theory and computerized adaptive testing.* New York: Academic Press.

Wernimont, G. (1977, March). Ruggedness evaluation of test procedures. *ASTM Standardization News*, *5*, 13–16.

Wernimont, G. (1978, December). Careful intralaboratory study must come first. *ASTM Standardization News*, *6*, 11–12.

Wheeler, J.A., & Zurek, W.H. (Eds.). (1983). *Quantum theory and measurement.* Princeton, NJ: Princeton University Press.

Widmalm, S. (1995, May 12). On exactitude, a review of *The Values of Precision*, by M. Norton Wise. *Science*, *268*, 905–6.

Wilson, T.P. (1971). Critique of ordinal variables. *Social Forces*, *49*, 432–444.

Wimsatt, W.C. (1981). Robustness, reliability and overdetermination. In M.B. Brewer & B.E. Collins (Eds.), *Scientific inquiry and the social sciences* (pp. 124–163). San Francisco, CA: Jossey-Bass.

Wise, M.N. (Ed.). (1995). *The values of precision.* Princeton, NJ: Princeton University Press.

Wolfe, E.W. (2000). Equating and item banking with the Rasch model. *Journal of Applied Measurement*, *1*(4), 409–434.

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

Woodcock, R.W. (1973). *Woodcock Reading Mastery Tests*. Circle Pines, MN: American Guidance Service.

Woodcock, R.W. (1999). What can Rasch-based scores convey about a person's test performance? In S.E. Embretson & S.L. Hershberger (Eds.), *The new rules of measurement: What every psychologist and educator should know* (pp. 105–127). Hillsdale, NJ: Erlbaum.

Wright, B.D. (1968). Sample-free test calibration and person measurement. In *Proceedings of the 1967 Invitational Conference on Testing Problems* (pp. 85–101). Princeton, NJ: Educational Testing Service.

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

Wright, B.D. (1984). Despair and hope for educational measurement. *Contemporary Education Review*, *3*(1), 281–288.

Wright, B.D. (1985). Additivity in psychological measurement. In E. Roskam (Ed.), *Measurement and personality assessment* (pp. 101–111). Amsterdam: North-Holland.

Wright, B.D. (1999). Fundamental measurement for psychology. In S.E. Embretson & S.L. Hershberger (Eds.), *The new rules of measurement: What every educator and psychologist should know* (pp. 65–104). Hillsdale, NJ: Erlbaum.

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

Wright, B.D., & Douglas, G.A. (1975). *Best test design and self-tailored testing*. (Research Memorandum No. 19). MESA Laboratory, Department of Education, University of Chicago.

Wright, B.D., & Linacre, J.M. (2003). *A user's guide to WINSTEPS Rasch-Model computer program, v. 3.38.* Chicago, IL: MESA Press [http::/www.rasch.org].

Wright, B.D., & Masters, G.N. (1982). *Rating scale analysis: Rasch measurement.* Chicago, IL: MESA Press.

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

Wright, B.D., & Stone, M.H. (1979). *Best test design: Rasch measurement.* Chicago, IL: MESA Press.

Wu, M.L., Adams, R.J., & Wilson, M.R. (1998). *ConQuest: Generalized item response modelling software.* Camberwell, Victoria: Australian Council for Educational Research.

ACKNOWLEDGEMENTS. Portions of this paper have been presented at various conferences over several years, including the Midwest Objective Measurement Seminar at the University of Chicago in 1989; the African Studies Association annual meeting in Baltimore, in 1990; the Eighth International Objective Measurement Workshops held at the University of California, Berkeley, in April 1995; and the American Psychological Association annual meeting in Boston, in 1999. Thanks to Benjamin Wright for his provocative dialogue; Mark Wilson, Willis Overton and Theo Dawson for their critical comments; Nathan Markward for sharing his energy and ideas; Henderikus Stam, Graham Richards, Joel Michell and an anonymous reviewer for their constructive suggestions; and to Richard Harvey, Robert Marier, Bronya Keats, Elizabeth Fontham and Jack Stenner for their support of this work over the course of the years.

WILLIAM P. FISHER, JR., is a Senior Scientist with MetaMetrics, Inc., of Durham, North Carolina, USA, and is widely published in the area of Rasch measurement theory and practice. He serves on the Editorial Board of the *Journal of Applied Measurement*, and on the Board of Directors of the Institute for Objective Measurement (http://www.rasch.org) and is a member of staff at the National Center for Special Education Accountability Monitoring. ADDRESS: MetaMetrics, Inc., 2327 Englert Drive, Suite 300, Durham, NC 27713, USA. [email: wfisher@lexile.com]