

PAMELA A. MOSS

## RECOVERING A DIALECTICAL VIEW OF RATIONALITY\*

**ABSTRACT.** In this article, I argue for an interpretation of Messick's (1989) theory of validity that supports a dialectical over a technical view of rationality in making validity judgments. A primary theme underlying Messick's theory is the "Singerian" approach to inquiry where one system of inquiry is observed by another in order to open "their underlying scientific and value assumptions to public scrutiny and critique" (pp. 61–62). Against Markus (this issue), who argues that a "completion" of Messick's theoretical project is necessary to support a single, best justified validity judgment for any given test use, I argue that Messick has provided a means of maintaining validity theory and the judgments it supports as ongoing accomplishments, always open to other perspectives, and critically reflexive in light of those challenges.

**KEY WORDS:** Messick, validity

*There is no surer path to awareness of unspoken preconceptions and prejudgments than communicative encounters with others who do not share them*

(McCarthy, 1994, p. 92)

In this response, I offer a different interpretation of Messick's (1989) theory of validity from the one proposed by Markus (this issue) – an interpretation that points in a substantially different direction for the further development of validity theory. Markus's reading, I will argue, is derived from and points towards a technical view of rationality where the goal is to develop inquiry procedures that can unambiguously adjudicate among competing interpretations or theories. In contrast, I will argue from and for a more dialectical view of rationality, one that supports ongoing critical reflection about our interpretations and theories in light of challenges from alternative perspectives. In drawing this contrast, I don't mean to suggest that those of us who develop and use tests or other social indicators do not regularly make practical decisions based upon our best judgment of the available evidence. Nor do I mean to



*Social Indicators Research* **45**: 55–67, 1998.

© 1998 *Kluwer Academic Publishers*. Printed in the Netherlands.

suggest that Markus's perspective precludes the critical practices and perspectives I have in mind. The issue is not really about what's *possible* within different perspectives (as Bernstein, 1979 notes), it's about what's emphasized, illuminated, or made more likely; what's relegated to the background as unimportant or impractical; and what the impact of these prevailing emphases is on the actual practices of social scientists and the communities they study and serve.

Markus argues that in order to achieve a unified theory of validity, a synthesis must be completed between Messick's evidential and consequential bases of validity. For Messick, the evidential basis encompasses construct validity plus evidence of the relevance and utility of scores to the applied purpose and setting; the consequential basis encompasses additionally the value implications and the social consequences of test interpretation and use. Markus characterizes Messick's evidential basis of validity as "value independent" and implying "one validity" for a given interpretation and use; whereas he characterizes Messick's consequential basis of validity as "value dependent" and suggesting "different validities for different sets of values" (p. 14). The synthesis Markus calls for entails the development of a theory of value justification that can result in a single, best justified value perspective and hence in a single validity for a given interpretation or use.

Underlying this argument are a set of assumptions Markus makes, with which I take issue, about aspects of Messick's theory. First is the assumption that the values implied in the consequential basis of test interpretation and use are distinct from the facts that would be generated in the evidential basis. Second is the assumption that values, as Messick conceives them, are and remain pre-rational and unjustified. Third is the assumption that Messick's Singerian approach to inquiry is analogous to Foucault's archetypal symbol of disciplinary technology, the Panopticon, and is intended to bring inquiring systems that "get out of hand ... back on line" (p. 8). Fourth is the assumption that a unified theory of validity must result in a single validity for a given interpretation and use. In this response, I will describe and then confront each of these assumptions, in turn, with my own reading of Messick. The thrust of my argument will be to highlight the dialectical view of rationality that I believe is explicit in Messick's work – a view which supports

validity theory as an always ongoing accomplishment, open to the challenge of new perspectives, rather than as a project to be completed.

I should note that this plan does not exhaust the detailed substantive argument in Markus's text nor my friendly disagreements with him. It does, however, emphasize those points that have, I believe, the most important implications for subsequent work in validity theory; and it permits a focused and coherent response in a limited number of pages. I should also note that I am not as confident of my reading of Markus as I would like to be. There are places in his text where I cannot tell whether he is offering his own interpretation of Messick or speculating about what reasonable readers might conclude. There are other places where Markus has taken seemingly different stances that I haven't yet been able to reconcile into a coherent interpretation of his perspective. I welcome his response and the opportunity it provides for him to contrast my reading of his text with his own.

#### IS THERE A "CORRECT" INTERPRETATION OF MESSICK?

Before undertaking the main task of challenging Markus's assumptions about Messick's text, I want to begin by explaining how I think two careful and respectful readers of Messick can come to such different conclusions. I will do so by sharing my own "philosophical conceits" about the nature of the dialogue in which Markus and I (and our readers and reviewers) are engaged, and second, to preview the dialectical view of rationality for and from which I argue.

The philosophers of social science whose work I (currently) find most persuasive would argue that there is no single correct interpretation that Markus or I can be seeking to approximate or against which our interpretations can be unambiguously evaluated (e.g., Gadamer, 1987; Greene, 1994; Hoy and McCarthy, 1994; Kogler, 1996; Taylor, 1987; Thompson, 1990). While some have argued (e.g., Hirsch, 1976) that the correct interpretation can be defined in terms of the original intent of the author, and that the role of interpreters is to bracket their prejudgments so as to reconstruct this intended meaning, others argue that this argument misunderstands understanding. Gadamer (1987), for instance, argues that there is

no knowledge without foreknowledge – without preconceptions or prejudices. “Understanding always implies a preunderstanding which is in turn prefigured by the determinate tradition in which the interpreter lives and that shapes his prejudices” (Gadamer, 1987: p. 87). “The task is not to remove all such preconceptions, but to test them critically in the course of inquiry . . . to make the all important distinction between blind prejudices and ‘justified . . . [or enabling] prejudices that are productive of knowledge’ (Bernstein, 1985: p. 128). Gadamer draws the useful analogy to interpretation in the reproductive arts (like the performance of a play): There is no single correct interpretation – every interpreter brings his own interpretation – but this is not an arbitrary interpretation that is independent from the original text, there is “a definable degree of appropriateness” (1987: p. 84).

This perspective does not imply that there are not better and worse interpretations, or that our interpretations cannot be comparatively evaluated, rather it acknowledges that our interpretations are necessarily contextualized and perspectival – that they involve a dialectic between the text itself and the foreknowledge/values on which we draw in interpreting it. Thus, understanding *and learning* progress as our foreknowledge/prejudgments are revised through respectful confrontation with others and with what we seek to interpret.

In the sections that follow, I describe and respectfully challenge (some of) the assumptions that Markus appears to make about Messick’s theory. In doing so, I question Markus’s call for a theory of value justification that will result in a *single* best justified validity along with his characterization of a unitary theory of validity.

#### ASSUMPTION 1: FACTS ARE DISTINCT FROM VALUES

Markus describes Messick’s framework as if facts were distinct from values: “The evidential basis [EB] is predicated upon a notion of rationality as something independent of values (Apel, 1979). Conversely the consequential basis [CB] is predicated on the diversity of social values” (Markus, p. 12). “The tension between EB and CB is therefore rooted in the division between facts and values” (Markus, p. 25).

TABLE I  
[Messick's] facets of validity

	Test interpretation	Test use
Evidential basis	Construct validity	Construct validity + relevance/utility
Consequential basis	Value implications	Social consequences

*Note:* In subsequent articles (e.g., Messick, 1994), the “progressive” nature of the matrix is illustrated by having the contents of each cell include the contents of all previous cells. So, for instance, construct validity appears in every cell.

Messick characterizes validity in a two dimensional table that distinguishes between interpretation and use on the one hand and the evidential and consequential bases on the other. The evidential basis subsumes construct validity plus the relevance and utility of test scores for the applied purpose and setting; the consequential basis additionally subsumes the value implications and social consequences of test interpretation and use.

The issue underlying the differences between Markus and myself has to do with how one interprets the categories and the lines represented in Messick's four-fold “facets of validity” matrix. Do they represent distinct components of the concept which can be considered separately from the others? Or do they simply name and illuminate aspects of the concept so that they will not be lost from our explicit consideration?

While I agree that the presentation of Messick's facets of validity *can* be read to support Markus's assumption about the distinction between facts and values; I believe the preponderance of evidence in the larger text points to an integrated view of facts and values. Consider the following:

Value implications are not ancillary but, rather, integral to score meaning. . . . These distinctions may seem fuzzy because they are not only interlinked but overlapping. (Messick, p. 20)

Elsewhere, Messick refers to the “the form-giving roles of values in *determining or distorting* the meaning of score interpretations per se” (Messick, p. 59), and the way in which “broader ideologies about the nature of humankind, society, and science . . . color our

manner of perceiving and proceeding” (p. 59). He notes that “data and values are intertwined in the conception of interpretation. And this applies not just to evaluative interpretation, where the role of values is often explicit, but also to theoretical interpretation more generally, where values assumptions frequently lurk unexamined” (p. 16). It thus appears that the evidential and consequential bases of validity cannot be distinguished in terms of facts and values.

ASSUMPTION 2: VALUES ARE PRE-RATIONAL  
AND WITHOUT JUSTIFICATION

Markus characterizes Messick’s view of values as being subjective – pre-rational and without justification – and then raises the concern that this subjective view of values makes it impossible to reconcile the evidential and consequential bases: “The assumption that facts are objective and values are subjective goes unquestioned in Messick’s theory . . .” (Markus, 1998: p. 2) “The resolution of the incomplete synthesis requires a reappraisal of the assumption that values lack rational justifications” (Markus, 1998: p. 4).

While Messick clearly addresses the ways in which values can “lurk” beneath the surface of our awareness, he argues that one of the goals of validity research is to illuminate and rationally evaluate those values: “emphasis is also given to the need for empirical and rational grounding of the value aspects as well as the substantive aspects of construct meaning” (p. 59). That is the express purpose of his Singerian approach to inquiry to which we will return in a moment.

Labeling scientific judgments as value judgements, as opposed to factual or theoretical judgments, does not absolve them from the need to be supported empirically or to be rationally defended against criticism . . . The intent is to illuminate the scientific and value assumptions of constructs and theories so that they may be subjected to either empirical grounding or policy debate, or both. (Messick, 1989: p. 63)

There is little question that Messick believes value perspectives can and should be justified. The only question is whether that justification must result in a uniquely or best justified evaluation. That is the issue to which I turn in the following sections.

ASSUMPTION 3: THE SINGERIAN APPROACH TO INQUIRY IS  
INTENDED TO BRING INQUIRING SYSTEMS THAT HAVE  
GOTTEN “OUT OF HAND ... BACK INTO LINE”

One of the primary themes that underlies Messick’s validity theory is represented in the Singerian approach to inquiry (Churchman, 1971; Singer, 1985, in Messick, in 1989) where one method of inquiry is evaluated in terms of another to highlight the assumptions and values underlying each. As Messick describes it:

A Singerian inquiring system starts with the set of other inquiring systems ... and applies any system recursively to another system, including itself. The intent is to elucidate the distinctive technical and value assumptions underlying each system application and to integrate the scientific and ethical implications of the inquiry. (p. 32)

Markus analogizes Messick’s Singerian inquiry to Foucault’s paradigmatic symbol of disciplinary technology, the panopticon: “The idea is to arrange the individual systems of inquiry into a scientific panopticon (Foucault, 1979/1975) in which systems each observe one another. If one gets out of hand, the other is there to bring it back into line” (p. 8).

For those who are unfamiliar with Foucault’s work, the Panopticon characterizes both a physical structure and a symbol of the means through which disciplinary technology exercises power over others (Foucault, 1977). Physically, the panopticon is an architectural plan intended to permit the monitoring and control of institutionalized populations. Dreyfus and Rabinow (1982) summarize Foucault’s characterization of its architectural features:

It consists of a large courtyard with a tower in the center and a set of buildings, divided into levels and cells, on the periphery. In each cell there are two windows: one brings in light and the other faces the tower, where large observatory windows allow for the surveillance of the cells ... The inmate is not only visible to the supervisor, he is *only* visible to the supervisor; he is cut off from any contact with those in adjoining cells ... The inmate cannot see if the guardian is in the tower or not, so he must behave as if surveillance is constant, unending, and total. (p. 189)

Foucault analogizes this physical symbol to multiple disciplinary technologies (such as those operative in prisons, factories, military institutions, hospitals, and schools) – any place where hierarchical observations (e.g., through examinations) and normalizing judgment locates people on a grid on which they can be compared. The

effect of panoptic technology is manipulation and control where others, perceiving themselves to be under constant surveillance, participate in their own domination and “normalization.”

Analogizing Singerian inquiry to panoptic technology reflects, I believe, an unfortunate reading of Messick – unfortunate in the sense that it undermines the potential of his theory to promote ongoing critical reflection and evolution in light of (sought-after) challenges from alternative perspectives. The point of Singerian inquiry is not to bring inquiring systems that have “gotten out of hand . . . back into line” (Markus, p. 8). Rather the point is to illuminate taken for granted practices and perspectives so that they can be critically evaluated by all involved:

It is precisely such mutual confrontation of theoretical systems, especially in attempting to account for the same data, that opens their underlying scientific and value assumptions to public scrutiny and critique. (Messick, pp. 61–62)

If I had to analogize Messick’s Singerian inquiry to an aspect of Foucault’s work, I would analogize its purpose to the purpose of Foucault’s genealogies. One role of genealogical inquiry is to highlight the “disqualified” knowledges “against the claims of a unitary body of theory which would filter, hierarchise and order them in the name of some true knowledge and some arbitrary idea of what constitutes science and its objects” (Foucault, 1980: p. 82). It is “through the reappearance of this knowledge . . . that criticism performs its work” (Foucault, 1980: p. 82). This emphasis on the importance of an outside perspective to illuminate what is taken for granted (as natural, normal, the ‘way things are done’) and thereby to provoke critical self-reflection is a theme that resonates across multiple philosophies of social science (e.g., Bernstein, 1985, 1992; Gadamer, 1979; Habermas, 1990; Greene, 1994; Hoy and McCarthy, 1994; Kogler, 1996) It is, in my judgment, one of the most profoundly important insights that Messick has brought to the tradition of educational and psychological measurement.

#### ASSUMPTION 4: A UNITARY THEORY OF VALIDITY REQUIRES A SINGLE VALIDITY ASSESSMENT

Markus’s argument rests, I believe, on the assumption that a unitary theory of validity requires a single uniquely or best justified validity



assessment: “The theory is presented as a unified theory of validity. That is ‘validity’ in the singular ... The unified emphasis is on combining various lines of evidence for this validity” (p. 13). The problem he constructs is to find a way to “reconcile the plurality of social values [implied in Messick’s consequential basis of validity] with the singularity of validity” (p. 13). The solution he proposes is to develop a theory of value justification that will result in a single validity assessment through one of three outcomes: (a) a uniquely justified validity assessment, (b) a validity assessment that is “indexed to specific sets of factual contingencies” (p. 26) and therefore uniquely justified for each contingency, or (c) a validity assessment that can be “ordered in a gradient of justification” and “stand out as most justified” (p. 27). Markus sees the completion of this synthesis, brought about through an appropriate theory of value justification that produces a best justified value, as “central to achieving a unified theory of validity” (p. 4).

In my judgment, defining a unitary theory of validity as one that results in a single validity assessment goes beyond what Messick means by the concept. Messick defines validity as “an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment” (p. 13). With respect to the unitary nature of the concept, he notes that “Although there are different sources and mixes of evidence for supporting scored based inferences, validity is a unitary concept. Validity always refers to the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores” (p. 13). This characterization is consistent with Markus’s emphasis on combining multiple lines of evidence, but it stops short of calling explicitly for a single (uniquely or best justified) outcome, and so it’s not yet clear whether Messick would agree with Markus’s interpretation.

A bit of history might help here. The earliest testing standards (APA, 1954, 1966) characterized different types of validity (e.g., content validity, construct validity, criterion-related validity) which were associated with different types of inferences – from the test score to a content domain, from the test score to a criterion variable,

or from the test score to a psychological construct that could not be defined by a content domain or a criterion variable. Since then, these “types” have come to be understood as types of evidence supporting a unified notion of validity. This is reflected in the 1985 *Standards’* use of the term content-, construct-, and criterion-related *evidence* in characterizing its “unitary concept” of validity. As Messick notes, “The varieties of evidence are not alternatives but rather supplements to one another. This is the main reason that validity is now recognized as a unitary concept” (p. 60).

Messick is careful to note that validity judgments are on-going accomplishments and validation an evolving process, always open to new evidence and theories.

It is important to note that validity is a matter of degree, not all or none. Furthermore, over time, the existing validity evidence becomes enhanced (or contravened) by new findings, and projections of potential social consequences of testing become transformed by evidence of actual consequences and by changing social conditions. Inevitably, then, validity is an evolving property and validation is a continuing process. Because evidence is always incomplete, validation is essentially a matter of making the most reasonable case to guide both current use of the test and current research to advance understanding of what test scores mean. (Messick p. 13)

Practically speaking then, validation is matter of making the most reasonable case with the available evidence while recognizing that these judgments will evolve as new evidence is brought to bear.

Elsewhere in Messick’s text, there are statements that appear to more directly challenge the presumption that validity theory must support convergence on a single outcome. Consider the following passages:

This [striving for consistency between trait implications and the evaluative implications] could prove difficult, however, because some traits, such as self-control and self-expression, are open to conflicting value interpretations. These cases may call for systematic examination of counterhypotheses about value outcomes – *if not to reach convergence on an interpretation, at least to clarify the basis of the conflict* (Messick, p. 60; italics mine).

The very recognition of alternative perspectives about the social values to be served, about the criteria to be enhanced, or about the standards to be achieved should be salutary in its own right. This is so because to the extent that alternative perspectives are perceived as legitimate, it is less likely that any one of these perspectives will dominate our assumptions, our methodologies, or our thinking about the validation of test use. (Messick, p. 88)

While acknowledging the practical reality of needing to make the best judgments we can with the available evidence, Messick also highlights the ever-present obligation to challenge those judgments with new evidence and theory and acknowledges the possibility that different perspectives may not converge.

#### CONCLUDING COMMENTS

Markus writes as if Messick has provided us with a vision of the end of the philosophy of social science, at least as it informs validity theory – “the reader may want to reflect on the success of this narrative in suppressing any doubts that we are indeed at the end of history” (p. 5) – and that if we can complete the difficult work of synthesizing the evidential and consequential bases of validity, we will have achieved it. The theorists I read would argue that any such perception of “completion,” consistent with a technical view of rationality, is dangerous: it risks narrowing “the range of the rational to what can be known with specific methods, thereby excluding whatever does not fit into this domain” (Hoy, p. 125). As McCarthy argues: “Intersubjective recognition of truth claims has to be *on-goingly accomplished* through rationally persuading one audience after another that it is ‘reasonable’ to accept them, that is, that there are good reasons for doing so, better reasons than for accepting any of the available alternatives” (McCarthy, p. 76). It’s true that within a more technical view of rationality “supposedly true facts can always be reexamined” (Hoy, p. 123), but what is missing from that picture is the illumination of taken-for-granted assumptions, values, and practices that alternative perspectives can provoke.

Messick’s Singerian approach to inquiry is a powerful means of maintaining the openness of validity theory – one that has tremendous potential for pushing our field forward. It opens our work not only to the voices that Messick has engaged in his chapter, those voices which, as he said, can be used to rationalize current validity thinking (p. 22), but also to all the voices we have yet to encounter or engage within our profession. It serves both a critical and a generative function, by simultaneously encouraging us to consider alternative methodologies and illuminating our own taken for granted practices for critique.

I know those of us engaged in the practical work of the development, evaluation, and use of social indicators will find such a perspective daunting. As I said at the outset, a dialectical view of rationality doesn't preclude us from making the best practical judgments we can with the information we have. It does, however, encourage us to be more cautious about the meaningfulness and the consequences of those judgments and it places a burden on the profession at large to support a critically reflexive research enterprise.

And so, I believe Messick has given us something far more valuable than a vision of a happy ending to the history of the philosophy of social science. Rather, he's given us the means of maintaining validity theory as an ongoing accomplishment, always open to other perspectives, and critically reflexive in light of those challenges. As Bernstein (1992) reminds us, "this kind of pluralistic dialogue is the responsibility of participants in any vital substantive tradition" (p. 66).

#### NOTE

\* Written in response to Keith A. Markus's article, "Science, Measurement, and Validity: Is Completion of Samuel Messick's Synthesis Possible?"

#### REFERENCES

- AERA, APA, and NCME: 1985, Standards for Educational and Psychological Testing (Authors, Washington DC).
- American Psychological Association: 1954, 'Technical recommendations for psychological tests and diagnostic techniques', [Supplement] Psychological Bulletin 51(2, Pt. 2).
- American Psychological Association: 1966, Standard for Education and Psychological Tests and Manuals (Author, Washington, DC).
- Bernstein, R. J.: 1992, *The New Constellation: The Ethical-Political Horizons of Modernity/Postmodernity* (The MPT Press, Cambridge).
- Bernstein, R. J.: 1985, *Beyond Objectivism and Relativism: Science, Hermeneutics, and Praxis* (The University of Pennsylvania Press, Philadelphia).
- Bernstein, R. J.: 1976, *The Restructuring of Social and Political Theory* (The University of Pennsylvania Press, Philadelphia).
- Dreyfus, H. L. and P. Rabinow: 1982, *Michel Foucault: Beyond Structuralism and Hermeneutics* (2nd ed.) (University of Chicago Press, Chicago).

- Foucault, M.: 1977, *Discipline and Punish* (A. Sheridan, Trans.) (Vintage Books, New York).
- Foucault, M.: 1980, 'Two lectures', in C. Gordon (ed.), *Power/Knowledge: Selected Interviews and Other Writings by Michel Foucault* (Pantheon Books, New York), pp. 78–108.
- Gadamer, H. G.: 1987, 'The Problem of historical consciousness', in P. Rabinow and W. M. Sullivan (eds.), *Interpretive Social Science* (University of California Press, Berkeley), pp. 82–140.
- Greene, M.: 1994, Epistemology and educational research: The influence of recent approaches to knowledge. *Review of Research in Education* 20, pp. 423–464.
- Habermas, J.: 1990/1980, 'The hermeneutic claim to university', in G. L. Ormiston and A. D. Schrift (eds.), *The Hermeneutic Tradition: From Ast to Ricoeur* (SUNY Press, Albany), pp. 243–272.
- Hirsch, E. D.: 1976, *The Aims of Interpretation* (University of Chicago Press, Chicago).
- Hoy, D. C.: 1994, 'Critical theory and critical history', in D. C. Hoy and T. McCarthy (eds.), *Critical Theory* (Blackwell, Oxford), pp. 101–214.
- McCarthy, T.: 1994, 'Philosophy and critical theory: A reprise', in D. C. Hoy and T. McCarthy (eds.), *Critical Theory* (Blackwell, Oxford), pp. 5–100.
- Kogler, H. H.: 1996, *The Power of Dialogue* (P. Hendrickson, Trans.) (The MIT Press, Cambridge, MA).
- Markus, K. A.: 1998 'Science measurement, and validity: Is completion of Samuel Messick's Synthesis Possible?'
- Messick, S.: 1994, 'The interplay of evidence and consequences in the validation of performance assessments', *Educational Researcher* 23(2), pp. 13–24.
- Messick, S.: 1989, 'Validity', *Educational Measurement* (3rd ed.) (The American Council on Education and the National Council on Measurement in Education, Washington, DC), pp. 13–103.
- Schmidt, L. K.: 1995, 'Introduction: Between certainty and relativism', in L. K. Schmidt (ed.), *The Spector of Relativism: Truth, Dialogue, and Phronesis in Philosophical Hermeneutics* (Northwestern University Press, Evanston, IL), pp. 1–22.
- Taylor, C.: 1987/1967, 'Interpretation and the science of man', in P. Rabinow and W. M. Sullivan (eds.), *Interpretive Social Science* (University of California Press, Berkeley), pp. 33–81.
- Thompson, J. B.: 1990, 'The methodology of interpretation', *Ideology and Modern Culture* (Stanford University Press, Stanford), pp. 272–327.

*4220 School of Education,  
University of Michigan,  
Ann Arbor, MI 48109-1259,  
U.S.A.*