JOURNAL OF PERSONALITY ASSESSMENT, *85*(2), 141–145

Rethinking Psychosocial Constructs: Reply to Comments by Barrett, Kagan, and Maraun and Peters

Robert E. McGrath
Department of Psychology
Fairleigh Dickinson University

In this article, I respond to comments provided by Barrett (2005/this issue), Kagan (2005/this issue), and Maraun and Peters (2005/this issue) in response to my article (McGrath, 2005/this issue). I agree with many of the points raised in their articles, although I disagree with some of Maraun and Peters's conclusions. Topics I address in this response include the meaning of measurement, the misapprehension that statistics can answer conceptual questions, problems with the concept of construct validity, and the problem of distinguishing between derived and fundamental measures.

One of the hardest tasks I faced in writing the article that begins this series (McGrath, 2005/this issue) was figuring out when to stop. Thinking about the limitations of the traditional psychometric model for scale development and what those limitations suggested about a potential alternative strategy raised a whole host of conceptual and practical issues for me to consider. To keep the task manageable, I had to make some judgments about which topics to pursue in detail, which merited a superficial look, and which were unimportant. Similar decisions had to be made about the content of the article, particularly as I wanted to produce something that was generally accessible. I must confess that the first draft I submitted was much longer than the one that was published. Thanks to some good feedback on that draft, I realized that the article would be greatly improved by sticking to the key issues, even if it meant that some topics near and dear to my heart received short shrift or were dropped completely.

So I am pleased by the articles that have been submitted in response to mine, not just because they raise some excellent points about conceptual and empirical issues in psychosocial measurement but also because they allow me the opportunity to revisit some themes that may have received less consideration than they should have in my article. I limit my comments largely to issues raised in the responses. In particular, because two of the articles focus primarily on fundamental questions about the nature and practice of measurement, my comments here are strongly biased to a more technical discussion of such issues.

CLARIFYING TERMINOLOGY

Looking at Barrett (2005/this issue) and Maraun and Peters (2005/this issue), one topic that probably merited more attention than I allotted to it was what I meant by *measurement* and related terms. The task of clarifying these terms is complicated by the fact that psychologists and philosophers of science have not been particularly consistent in their use.

References to measurement in my article were intended to refer to any operation that systematically applies numerals to entities or events as per the philosophy of measurement first introduced by Stevens (1946); I was not intending to equate measurement with quantitative measurement. In the sentence Barrett (2005/this issue) quotes from the start of my article, I really meant to suggest that systematic observation is a key component of science.

I propose this usage of the term knowing that it has been criticized for being overinclusive (see Michell, 1990). More restrictive definitions have been proposed. N. R. Campbell (1920/1957), for example, restricted measurement to quantitative measurement, defined in turn as scales with values capable of addition or derivation from scales that are capable of addition (i.e., interval or ratio scales). Others (e.g., Luce, 1986) permitted ordinal measurement. However, Stevens's (1946) formulation remains the most familiar to psychologists. It is also consistent with common language use of the term *measure,* potentially providing an example of Maraun

and Peters's (2005/this issue) claim that psychologists tend to confuse normal and formal concepts.

Consistent with this loose conceptualization of measurement, I tended to use the terms *measure* and *indicator* interchangeably. As Maraun and Peters (2005/this issue) point out, there is a worthwhile distinction between these two. I would suggest that an indicator be defined as a systematic assignment of numerals to observations without the assumption that an isomorphism exists between assigned values and placement on some construct. This is a definition Maraun and Peters might reject, but I use it to reformulate the problem with psychosocial measurement.

Many scientific disciplines rely at times on indicators as the basis for variables such as the index of coincident economic indicators in macroeconomics. An important difference is that almost all psychosocial variables are derived from indicators rather than direct measures, with behavioral frequency counts representing perhaps the only popular exception. This is not a desirable state of affairs. Indeed, the basic proposition of the article that begins this series could be restated as follows: Measurement based primarily on indicators rather than direct measures is a problematic basis for scientific progress.

SOME SUPPORTIVE ADDITIONS

In several cases, the commentators have done me a great service in providing more ammunition for some of the points I raised. I think Barrett's (2005/this issue) question "what if there were no psychometrics?" augments my central argument nicely. The blind quest for reliability is simply too weak a standard for adequate measurement. Without psychometrics, psychologists would not have been so successful at glossing over the vagueness inherent to psychosocial constructs, the arbitrary quality of psychosocial scales, and the tenuous relationship between those constructs and those scales.

It is worth noting that Barrett (2005/this issue) is not the first to question the psychometric assumptions that inspire much of measurement in psychology or the extent to which psychometrics has helped psychologists avoid reconsidering those assumptions (e.g., Michell, 2000, 2001). To quote a particularly important analysis of this issue (Krantz, Luce, Suppes, & Tversky, 1971)

> Most of the psychometric literature is based on numerical rather than qualitative relations (e.g. matrices of correlation coefficients, test profiles, choice probabilities) … Here by contrast, we are concerned almost exclusively with the qualitative conditions under which a particular representation holds. … For example, in scaling aptitude, intelligence, or social attitudes, test scores or numerical ratings are usually interpreted as measures of the attribute in question. But in the absence of a well-defined homomorphism between an empirical and a numerical relational structure, it is far from clear how to interpret such numbers. (pp. 32–33)

Unfortunately, Krantz et al.'s powerful indictment of current practice in measurement has largely been ignored in the psychosocial literature (Cliff, 1992). I can only hope that Barrett's challenge is taken as seriously as it deserves to be and receives the same level of attention as was awarded to the question when applied to significance testing.

Barrett (2005/this issue) goes beyond classical test theory and criticizes the use of item response theory (IRT) to clarify latent variables. I appreciate Barrett's choice of this topic, as IRT has in recent years been touted as a means of demonstrating that psychosocial measures meet minimum criteria for being considered quantitative measures (e.g., Barrett, 2003; Karabatsos, 2001). Barrett's conclusion is similar to the one I drew concerning standardization in my article, one that seems to apply equally well to all the statistical methods—including factor analysis and structural equation modeling—that have been touted over the years as a means for clarifying constructs: Statistics alone without clearly defined constructs cannot resolve conceptual issues.

Kagan (2005/this issue) delivers a compelling set of arguments against overly abstract constructs and the overreliance on self-reports. I am particularly pleased that Kagan chose to submit a response to the article, as his previous discussions of these issues (e.g., Kagan, 1988) were seminal to the development of my own ideas. Whether the suggestions I provide at the end of my article offer the basis for an adequate solution to the problem, Kagan (2005/this issue) makes it difficult to contend that the problem does not exist.

Finally, I would like to make an addition of my own, another problem with multi-item scales to add to those I described in the section titled "Multi-Item Scales as Representations" (McGrath, 2005/this issue). Statistical practice in psychology tends to treat the variables that result from multi-item scales as quantitative variables capable of addition. In fact, there is insufficient evidence even to ensure that these variables are ordinal. In the absence of a set of weights for items based on their objective relevance to the construct, it is probably impossible to prove that the resulting scale consistently orders individuals in a manner that can be considered accurate. Notice this inability to corroborate the ordinality of scale values does not result from response bias or unreliability; it is an inherent flaw in the multi-item scale as a representation of placement on a complex, loosely defined social construct.

CONCEPTS VERSUS CONSTRUCTS

The most challenging of the three responses is the article by Maraun and Peters (2005/this issue), as they disagree with several of the basic tenets underlying my analysis. I do not take Maraun and Peters's objections personally. Maraun and Peters clearly have larger fish to fry, namely, the philosophy of construct validation that has predominated in psychological measurement for the last 50 years. I start by saying I have sympathy

for Maraun and Peters's complaint that construct validity as commonly formulated confounds empirical and conceptual issues. I decided to couch my article in terms of construct validity to provide a context most psychologists could understand, but I was not completely comfortable with this decision. It seemed to me that Cronbach and Meehl (1955) linked issues of construct meaning (what Maraun & Peters, 2005/this issue, might refer to as concept formation) too closely with mapping of the nomological net (which Maraun & Peters, 2005/this issue, would correctly consider an empirical activity). In fact, the first draft of my article focused more extensively on Cone's (1995) distinction between the representational and elaborative validity of a scale. Although this is a different formulation than Maraun and Peters's, Cone was dealing with the same lack of clarity in the concept of construct validity. Statements such as "scientifically speaking, to 'make clear what something is' means to set forth the laws in which it occurs" (Cronbach & Meehl, 1955, p. 290), and "a construct is defined implicitly by a network of associations or propositions in which it occurs" (Cronbach & Meehl, 1955, pp. 299–300), obscure an important distinction between what a construct is and how well the scale reflects that versus the nomological net of the construct or the pattern of correlations for the scale. I hope it is clear in my article that when I refer to construct validity, I am more interested in the former than in the latter.

At the same time, I do not want to draw as fine a line between conceptual and empirical activities as Maraun and Peters (2005/this issue) drew for several reasons. First, doing so requires a fairly radical shift in thinking about abstractions in psychology. Where concepts reflect Wittgenstein's (1953) linguistic idealism, most psychologists tend to understand psychosocial abstractions in terms of some variant of critical realism (D. T. Campbell, 1995; Manicas & Secord, 1983). The definition of scientific constructs and the methods for measuring them are treated as completely embedded within empirical activities that attempt to understand the "true" nature of those constructs (a philosophical position that I do not think is incompatible with viewing complex psychosocial constructs as social constructions, as I indicated in my article). These empirical activities are not restricted to the description of attributes characteristic of natural processes or objects, or of their relationships with other processes and objects, as Maraun and Peters suggest; they are integral to the process of developing and refining the necessary and sufficient conditions for concluding that a construct denotes a certain object. For example, the only instance Maraun and Peters (2005/this issue) provide of a "proper technical concept" (p. 129) in psychology, premsia, is difficult if not impossible to understand outside the context of the analytic strategy by which it was developed. For these reasons, I continue to refer to the realist's constructs rather than the idealist's concepts.[1]

The sharp distinction Maraun and Peters (2005/this issue) draw between empirical and conceptual activities is a fairly radical departure. Maraun (1998) rejected both construct validation (Cronbach & Meehl, 1955) and Krantz et al.'s (1971) representational theory of measurement, two of the best known approaches to understanding measurement in psychology today. This divergence from the mainstream of psychology even emerges in comparisons with the other two commentaries. Maraun and Peters conclude it is absurd to question the validity of a conceptualization, yet Barrett (2005/this issue) considers the validity of conceptualizations "the heart of the issue" (p. 136). Maraun and Peters conclude my characterization of complex psychological constructs as hierarchical is inaccurate, but in fact, Barrett (2005/this issue) assumes the accuracy of hierarchical structuring in his discussion of his Figure 1 (see also Paunonen, 1998), and Kelly's (1955) personal construct theory assumes constructs are hierarchically ordered.[2] Although Maraun and Peters deny that natural language concepts relevant to psychology include an assumption of cross-modal consistency, this consistency assumption is strong enough that Kagan (2005/this issue) offers several instances in which he seems to believe psychologists would find the results surprising because of the absence of such consistency (see also D. T. Campbell & Fiske, 1959). By drawing too fine a distinction between conceptual and empirical activities, Maraun and Peters are incorrect about the way in which psychologists' understanding of psychosocial constructs is formed by the methods of domain sampling. I reject the conclusion that just because psychologists do not have a tradition of offering technical definitions for their constructs, they mean the same thing as the lay person when they use a term.

Finally, I am not clear what Maraun and Peters (2005/this issue) are suggesting we do instead, and so I am hesitant to accept such a full-scale denunciation of traditional conceptions at this time. Maraun (1998) seemed quite pessimistic about the potential for a solution at all. I agree wholeheartedly with Maraun and Peters's (2005/this issue) assertion that

> The chief concern of the psychologist has been to investigate psychological phenomena of interest to humanity, and these phenomena are precisely those that are denoted by concepts that are a part of ordinary language. … However, the correct employments of ordinary language concept[s] … are not

[1]Maraun and Peters (2005/this issue) indicate that construct validation is based on "initial misunderstandings of the empirical realist philosophy" (p. 131), so I may well be misrepresenting the extent to which they have rejected realism for idealism. They unfortunately do not expand on this point; I look forward to reading more from them on this issue.

[2]In an interesting parallel to Maraun and Peters's (2005/this issue) claim that complex constructs are more accurately described in terms of "complex grammars" (p. 130) rather than in terms of hierarchies, several commentaries have questioned the accuracy of Kelly's (1955) model of hierarchical organization (e.g., Husain, 1983), so the hierarchical assumption probably deserves further consideration despite its popularity among psychologists.

> fixed by necessary and sufficient conditions but rather by a bewildering variety of other types of rule. … Psychological concepts have complex, unwieldy grammars that are "widely ramifying, lacking in unifying employment and not readily surveyable." (pp. 129)

I am uncertain whether, as Maraun and Peters seem to be suggesting, the problem can be resolved by defining psychological constructs in terms of (necessary and sufficient?) behavioral criteria and developing scales that consist of those behavioral criteria. If my understanding of their solution is correct,[3] it seems unlikely to me that you can define a construct such as anxiety in terms of behavioral criteria without reducing its social usefulness (which, as Maraun and Peters point out, was the undoing of premsia) or facing the practical problems of multi-item scaling described in my (McGrath, 2005/this issue) article.[4]

FUNDAMENTAL AND DERIVED MEASUREMENT

That being said, I cannot at this point consider the solution I offer in my article (McGrath, 2005/this issue) any more certain of success. The immediate problem is this. I, Barrett (2005/this issue), Kagan (2005/this issue), and Maraun and Peters (2005/this issue) all seem agreed that psychology can only advance so far without clarity in the constructs we conceive. I, Kagan, and perhaps Barrett also seem agreed that the most likely option for achieving this goal would involve focusing on more fundamental, elemental, unidimensional constructs. To do so, though, there must be some strategy for distinguishing elemental constructs from those that are synthetic, abstract, and complex. The differentiation of fundamental from derived measurements is a longstanding problem even in quantitative measurement (see Krantz et al., 1971), one for which unfortunately there is no simple solution. Furthermore, I suspect Kagan is correct in that the lexicon of fundamental constructs will vary across modes of measurement, and the key constructs within modes will vary across cultures even if the set of constructs does not. If in fact the solution to imprecision in measurement lies in the study of elemental rather than complex constructs, the first task to be addressed in a program of research on this topic is the development of techniques for isolating fundamental psychosocial constructs, a task that I suspect will require input from both personologists and members of the populations targeted for measurement. There is more to follow.

[3] I accept the possibility it might not be because their description of this proposal is unfortunately short on details.

[4] I would also like to clarify an issue raised by Maraun and Peters (2005/this issue) in their footnote 3. Maraun and Peters draw a conclusion from a statement in my article that was unintended. In stating "prediction occurs in the context of operationism in that the goal of successful measurement is solely the maximization of an observable relationship" (McGrath, 2005/this issue, p. 120), it was not my intention to caricature operationism as a philosophy of maximizing relationships. I hope my reference to the diagnostic nomenclature as an operationist system of constructs makes clear that I consider operationism much more than that. My contention was only that psychologists' use of statistical significance as a basis for deciding what is important was in part rationalized on the basis of operationism.

REFERENCES

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

Barrett, P. (2005/this issue). What if there were no psychometrics? Constructs, complexity, and measurement. *Journal of Personality Assessment, 85,* 134–140.

Campbell, D. T. (1995). The postpositivist, nonfoundational, hermeneutic epistemology exemplified in the works of Donald W. Fiske. In P. E. Shrout & S. T. Fiske (Eds.), *Personality research, methods, and theory: A Festschrift honoring Donald W. Fiske* (pp. 13–27). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81–105.

Campbell, N. R. (1957). *Foundations of science: The philosophy of theory and experiment.* New York: Dover. (Original work published 1920)

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

Cone, J. D. (1995). Assessment practice standards. In S. C. Hayes, V. M. Follette, R. M. Dawes, & K. Grady (Eds.), *Scientific standards for psychological practice: Issues and recommendations* (pp. 201–224). Reno, NV: Context.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302.

Husain, M. (1983). To what can one apply a construct? In J. R. Adams-Webber & J. C. Mancuso (Eds.), *Applications of personal construct theory* (pp. 11–28). Toronto, Ontario, Canada: Academic.

Kagan, J. (1988). The meanings of personality predicates. *American Psychologist, 43,* 614–620.

Kagan, J. (2005/this issue). A time for specificity. *Journal of Personality Assessment, 85,* 125–127.

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

Kelly, G. A. (1955) *The psychology of personal constructs.* New York: Norton.

Krantz, D. G., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement* (Vol. 1). New York: Academic.

Luce, R. D. (1986). Uniqueness and homogeneity of ordered relational structures. *Journal of Mathematical Psychology, 30,* 391–415.

Manicas, P. T., & Secord, P. F. (1983). Implications for psychology of the new philosophy of science. *American Psychologist, 33,* 399–413.

Maraun, M. D. (1998). Measurement as a normative practice: Implications of Wittgenstein's philosophy for measurement in psychology. *Theory and Psychology, 8,* 435–461.

Maraun, M. D., & Peters, J. (2005/this issue). What does it mean that an issue is conceptual in nature? *Journal of Personality Assessment, 85,* 128–133.

McGrath, R. E. (2005/this issue). Conceptual complexity and construct validity. *Journal of Personality Assessment, 85,* 112–124.

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

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

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

Paunonen, S. V. (1998). Hierarchical organization of personality and prediction of behavior. *Journal of Personality and Social Psychology, 74,* 538–556.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103,* 667–680.

Wittgenstein, L. (1953). *Philosophical investigations.* Oxford, England: Blackwell.

Robert McGrath
School of Psychology
T–WH1–01
Fairleigh Dickinson University
Teaneck, NJ 07666
Email: mcgrath@fdu.edu

Received March 11, 2005