

Book Reviews

Borsboom, D. (2005). *Measuring the mind: Conceptual issues in contemporary psychometrics*. Cambridge, UK: Cambridge University Press.

Reviewed by
Jacqueline P. Leighton
University of Alberta

Once in a while a book comes along that captures an idea, a zeitgeist, so convincingly and articulately, that you know before you finish that it has already enriched the way you think about a topic. These are the books that one finishes speedily as if devouring a good meal. It is not that the contents of the book are ideal, without some limitations, but rather, that the contents of the book reflect so well an idea whose time one hopes has come, thorny as it may be to implement in current practice. *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics* by Denny Borsboom represents such a book. This book represents a culmination of ideas published previously in journals such as *Psychological Review*, *Applied Psychological Measurement*, and *Intelligence*. In this review, I describe the six chapters of this 185-page book, which also includes references and index. I also note the limitation I see with its contents and how it contributes to educational measurement. I urge others to read it, especially those interested in cognitive diagnostic assessment and validity theory.

The six chapters of the book reflect an impressive interplay between philosophy of science, measurement, and mathematics. Consequently, readers who enjoy probing the *why* behind how we think about true scores, latent variables, scales, relations between models, and ultimately validity will, I think, relish the contents of the book. The first chapter introduces the reader to the question that originated the book—Do tests really measure something and, if so, what is it? Borsboom suggests that “after a century of theory and research on psychological test scores, for most test scores we still have no idea whether they really measure something, or are no more than relatively arbitrary summations of item responses” (p. 2). Although Borsboom uses psychological tests of intelligence and of personality as a basis for many of his examples, the reasoning behind his discussions are applicable to educational tests of academic achievement, domain mastery, competency certification, and cognitive skills. Borsboom indicates that one of the central ideas behind the book is “that too little attention has been given to a conceptual question about measurement in psychology: what does it *mean* for a psychological test to measure a psychological attribute?” (p. 3). The goal of the book then is to explore answers to this perplexing, but important conceptual question, and entailed consequences.

In order to begin to consider the set of possible answers to the conceptual question of what does it *mean* for a psychological test to measure a psychological attribute, Borsboom delves into classical test theory, latent variable theory, and

representational measurement theory in chapters 2, 3, and 4, respectively. In chapter 2, he takes the reader through the syntax (model formulation) and semantics (interpretation of the model formulation) of classical test theory. In doing so, he engages us in a thought experiment originally presented in Lazarsfeld (1959; also used by Lord & Novick, 1968) involving *Mr. Brown*, a fictional character that has his memory erased after he answers a question so that his responses can be considered independent. Following the syntax and semantics, Borsboom evaluates how the assumptions of the model are reconciled with data in actual, real-life situations such as in the calculation of test reliability. Finally, he examines the psychometric model from an ontological stance, which means that he scrutinizes what true scores actually mean, if anything, in relation to people and their psychological attributes. He concludes that classical test theory is ontologically ambiguous, confounding reliability and validity in true score interpretations.

A similar analysis is undertaken of latent variable models and scales (representational measurement theory) in chapters 3 and 4, respectively. In his analysis of latent variable models in chapter 3, Borsboom concludes that these models reflect a realist ontology—what is measured by the test may not be directly observed but exists covertly in some form. He considers latent variable models to be a vast improvement over classical test theory because a causal relation is conceptualized between the attribute to be measured and item responses. Borsboom explains that limitations also exist with latent variable models. In particular, it is problematic to defend causal statements relating psychological attributes to item responses especially in a “within-subject sense” because “an individual’s position on the latent variable is, in a standard measurement model, conceptualized as a constant, and a constant cannot be a cause” (p. 70). Of course, one could argue that the position on the latent variable is only viewed as a constant at the time of the test, but is not really a constant in real time because of its propensity to change with learning. However, this qualification may not necessarily facilitate the inference of causality because usually we have a datum on an individual’s position on the latent variable at the time of the test alone. Defending causal statements that relate psychological attributes to item responses in a “between-subject sense” is less problematic because at least there is covariation between the posited psychological attribute and the response—differences between subjects in their position on the latent variable precedes differences in expected item responses. Borsboom also presents an interesting discussion on *locally homogenous* and *locally heterogeneous* constructs. He suggests that there is a strong uniformity assumption that between-subject results will answer within-subject questions; for example, that a latent factor structure accounts for the processes all participants used to answer a series of test items. Borsboom indicates that it is not at all evident that psychological attributes should exhibit the same dimensionality (i.e., be locally homogenous) across these individuals. In chapter 4 on scales, Borsboom argues that while representational measurement theory has few metaphysical assumptions, a literal reading of the theory suggests that it has no clear means of incorporating error in observations. Unlike latent variable theory, representational measurement theory involves the assumption that what is being measured can be recreated without error. This is problematic because “what matters is not just the structure of the data, but also the question of how these data

originated" (p. 105). Thus, latent variable models may be as good as it gets for now especially for the kinds of psychological and educational measurement objectives of interest.

In chapter 5, Borsboom tackles the relations between the different mathematical models (e.g., in classical test theory, latent variable theory, and representational theory) for psychological and educational measurement. For example, he reiterates that "classical test theory is basically about the test score themselves, representationalism is about the conditions that should hold among test and person characteristics in order to admit a representation in the number system, and latent variable theory is about the question where the test scores come from" (p. 121). Despite the differences in philosophical underpinnings, Borsboom illustrates the mathematical relations among models at the syntactic level. However, he raises the question of how these models relate semantically and even ontologically to each other. Borsboom argues that if we can tolerate the metaphysics of propensities (i.e., physical characteristic of an object, dispositions to behave in a certain way) and interpret probabilities in these models as *propensities* of individuals at a specific point in time, all three models are complementary at the semantic and ontological levels.

The first five chapters of *Measuring the Mind* usher in perhaps the most interesting chapter of the entire book—the concept of validity. Given Borsboom's examination of three different models in terms of their syntax, semantics, and ontology, the grand finale is appropriate for what all this reveals about validity. What does it *mean* for a psychological test to measure a psychological attribute? Borsboom suggests that validity is a property of tests. And "a test is valid for measuring an attribute if and only if (a) the attribute exists, and (b) variations in the attribute causally produce variations in the outcomes of the measurement procedure" (p. 150). This view of validity carries with it substantial implications. For starters Borsboom argues "the crucial mistake is the view that validity is about correlation. [In fact,] [V]alidity concerns measurement, and measurement has a clear direction. The direction goes from the world to our instruments. It is very difficult not to construct this relation as causal" (p. 160). He explains that in the natural sciences, few researchers create measurement instruments without having a reasonable idea of the processes that lead to expected measurement outcomes. What this entails for those readers interested in truth-seeking questions about what it really *means* for a test to measure a psychological attribute is simply the following: "the primary objective of validation research is not to establish that the correlations go in the right directions, but to offer a theoretical explanation of the processes that lead up to the measurement outcomes" (p. 163).

Upon reflection, the main limitation of the book is that its core message has been delivered before, albeit not as forthrightly. Theorists and researchers (e.g., Embretson, 1983; Loevinger, 1957; Messick, 1989; and now Kane, 2006) have been advocating a better understanding the psychological processes we intent to measure for some time. Borsboom recognizes this legacy, but perhaps the repetition of important messages within different conceptual frameworks is a necessary evil that eventually persuades us all to move forward. Now what we need are innovative methods, illuminating studies, and clearer roadmaps to begin to seize this daunting challenge. To be sure, Borsboom's account is unequivocal with respect to the semantic and

ontological clarity to which measurement theory should aspire. Such force in exposition makes it difficult to dismiss.

References

Embretson, S. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin, 93*, 179–197.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Westport, CT: National Council on Measurement in Education and American Council on Education.

Lazarsfeld, P. F. (1959). *Latent structure analysis*. In S. Koch (Ed.), *Psychology: A study of a science*, vol. 3(pp. 476–543). New York: McGraw-Hill.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3*, 635–694 (Monograph Suppl. 9).

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: American Council on Education/Macmillan.

Author

JACQUELINE P. LEIGHTON is Associate Professor of Educational Psychology, Centre for Research in Applied Measurement and Evaluation (CRAME), University of Alberta, 6-110 Education North, Department of Educational Psychology, Faculty of Education, Edmonton, Alberta, Canada T6G 2G5. Her primary research interests include cognition and assessment.