

CAN WE BRING ABOUT A VELVET REVOLUTION IN PSYCHOLOGICAL
MEASUREMENT? A REJOINDER TO COMMENTARIES

DENNY BORSBOOM

UNIVERSITY OF AMSTERDAM

First, I thank Sijtsma, Clark, Kane, and Heiser for taking the time and effort to provide a commentary on my paper, and the Editor for allowing me to respond to them. In general, the commentators agree with the thesis that psychometrics and psychology are, to an extent that must be deemed problematic, disconnected. They further agree with the upshot of this diagnosis: Psychometricians need to work harder to make a difference in psychology, and psychologists need to develop a greater awareness of important psychometric developments. However, the commentators also raise several points of disagreement and criticism. I focus on four important topics that concern:

- (a) the theoretical factors influencing the current state of affairs;
- (b) the question of how to proceed in the absence of theory;
- (c) the communication between psychometricians and psychologists; and
- (d) the relation between experimental psychology and psychometrics.

Theoretical Factors Revisited

In an attempt to analyze the theoretical factors contributing to the present state of affairs, I identified as likely perpetrators: Operationalism, classical test theory, and construct validity. Kane, Sijtsma, and Heiser disagree with various aspects of this diagnosis.

Both Kane and Sijtsma argue that, in the absence of detailed formal theory, measurement models are necessarily driven by pragmatic factors and global considerations about the attribute of interest. Latent variables that occur in such models are therefore, at least initially, operationally defined. As such, operational definitions play a natural and important role in scientific progress. For this reason, to demand full-blown formal theoretical definitions *before* the first attempt at measurement is to put the cart before the horse. In an important sense “[a] specification of how a theoretical attribute produces certain effects is possible only after the theory is in place; it is not the first step, but one of the last steps in a research program” Kane (2006, p. 444). Sijtsma (2006, p. 453) makes a similar point in arguing that the specification of the causal relation between an attribute and the observations, as advocated in Borsboom, Mellenbergh, and Van Heerden (2004), while perhaps desirable, may be premature because “we still know so little about the functioning of the human brain in general and cognitive processes including those underlying personality traits and attitudes in particular.”

Thus, Kane and Sijtsma argue that an orienting ideal, that is, knowledge of how a test actually works, should not be categorically forced upon psychology as a necessary requirement for test use; researchers need breathing space to work with operational definitions, general conceptualizations

Requests for reprints should be sent to Denny Borsboom, Department of Psychology, Faculty of Social and Behavioral Sciences, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam. E-mail: d.borsboom@uva.nl

of attributes, and the liberal forms of construct validity theory. I agree with much of this—as long as everybody knows what they are doing and it is clear that we are proceeding in the right direction. Now, perhaps Kane and Sijtsma are in the good part of psychology, the one that makes all the quick and indisputable progress. In the part of psychology where I live, however, all too often researchers seem to have little sense of psychometric direction. They follow dubious guidelines, like “if alpha is higher than .80 the test is reliable,” “if the scree-plot shows a good drop the test measures only one thing,” and “if the test scores show some reasonable pattern of correlations the test must have some degree of validity”; Clark mentions and criticizes some similar practices. It seems to me that, in such cases, operational definitions, construct validity, and classical test theory are not functioning in the healthy way that the methodology books describe, but instead have become instances of dogmatic thought that hinder progress in psychology. Hence, while I recognize and acknowledge that many of the ideas in classical psychometrics have a legitimate role to play in scientific practice, as Kane, Heiser, and Sijtsma stress, at the same time I would urge psychometricians to take a closer look at how these concepts are functioning in actual scientific practice. I think they will find that much of psychology is not following the idealized template of scientific progress that underlies Kane’s and Sijtsma’s nuanced points on measurement, but is rather, quite simply, lost when it comes to dealing with psychometric questions. I also think that there is serious room for improvement, because, in contrast to fifty years ago, when the foundations of classical test theory and construct validity were laid down, we now have the computing power and mathematical tools to make modern psychometric models work.

That operationalism, construct validity, and classical test theory are not currently playing a productive role in steering psychology forward does not mean that the ideas and concepts in these theories are *generically* useless. Heiser rightly criticizes such a point of view with respect to classical test theory, but misattributes it to me. I actually think classical test theory is quite useful, for instance,

- (a) as a statistical theory of composite test scores;
- (b) as a first-order approximation to a restricted class of latent variable models (Holland & Hoskens, 2003); and
- (c) as a source of established theorems that can shed light on the relation between theoretical attributes, item response models, and observed test scores (Ellis & Junker, 1997).

However, classical test theory *is* useless as a *measurement model*. The reason is that measurement requires one to relate attributes to test scores, and classical test theory does not have the expressive power to do so (see Borsboom, 2005, pp. 32–33, for a detailed argument). If one is only interested in test scores, classical test theory is fine; but to deal with the measurement problem, one needs a latent variable model.

What To Do in the Absence of Theory?

Latent variable models are most useful if there exists a reasonably detailed view of the structure of attributes and the way that response processes connect this structure to the observations. Sijtsma and Kane both address the problem that this sort of theory is often lacking. Sijtsma (2006, p. 453), for instance, states that “[t]aking substantive theory as a starting point for test construction is an excellent idea that has existed for a long time but is not widely practiced. The reason probably is that much theory is still in its puberty, infancy, or even at the fetal stage.” Kane (2006, p. 453) states that “[a]ssuming that no formal theory exists (the usual case), the test-development process is necessarily ad hoc, guided by general conceptions of the attribute of interest.” For precisely this reason, I argued that it is important to develop substantive psychological theory as it relates to measurement, and suggested that psychometricians should play a more pronounced

role in this process. Without such theory, validation research will indeed be never-ending, as has been argued elsewhere (Borsboom, Mellenbergh, & Van Heerden, 2004), because it means that one is stuck with a black box model of the relation between attributes and test scores. Sijtsma (2006, p. 453), however, thinks this point of view is “impractical,” Kane appears to view the ad-hoc development of measurement instruments as something that we simply have to live with, and Heiser (2006, p. 459), states that psychologists “have to take a leap of faith anyhow” so that it does not really matter to which measurement model their leap leads.

This is a remarkable reaction to the current state of affairs. If we all agree that stronger substantive theory is needed to get a grip on the measurement problem, and such theory does not presently exist, then is it really the proper course of action to settle for second best, and just model whatever data substantive psychologists come up with, so that we basically assist them in fitting ad-hoc statistical models to ad-hoc constructed tests? Or should we attempt to develop the required theory? I take it to be rather preliminary to admit defeat at this early stage of theory development, and hence would argue that we take the second option seriously. And if we do this, then why should psychometricians not pay attention to the formalization and development of psychological theory insofar as it relates to measurement? Do psychometricians take the “psycho” in psychometrics seriously, or has psychometrics really become a subdivision of statistics, so that the subtitle of this journal should be changed from “a journal for quantitative psychology” to “a journal for data analysis”? I hope that, in this case, to ask the question is not to answer it.

Communication Breakdowns

One of the causes of the lack of integration between psychometrics and psychology undoubtedly lies in communication problems. Now, with respect to the communication breakdown that is currently in place, Clark reacts strongly to my suggestion that psychometric papers are often ignored in psychology, and lays the blame partly on psychometricians. In reference to Millsap’s work on measurement invariance, for instance, she states that “[i]f Millsap (1997) were written so I could assign it in a first-year graduate class, [it] would have a better chance to advance the field” (Clark, 2006, p. 449). I think Clark is correct here, but at the same time feel that this illustrates one of the major problems that hinder psychometric advances in psychology. Think of it again: *If one wants to communicate a psychometric advance, the paper has to be written at a level that allows it to be assigned to a first-year graduate class.* The reason that this is true is exactly what makes it food for thought. Are there no important ideas in psychometrics that cannot be understood by the average first-year graduate student?

Now, all this should not be taken to mean that psychometricians are somehow “superior” to psychologists, or that substantive psychologists “deserve to be mocked,” as Clark states (2006, p. 448). As far as I can see, we are all dwarfed by the problem of explaining human behavior; hence nobody is justified in taking a “superior” attitude—and it was certainly never my intention to do so. Further, I agree that it is important that psychometricians write accessible papers that are as clear as possible; and there certainly is much room for improvement on this score. However, I also think that, when the content of a psychometric paper is as manifestly important as Millsap (1997), some effort can reasonably be expected on the part of the reader.

But how important, exactly, *are* such psychometric insights as communicated in Millsap’s paper? Clark (2006, p. 449) expresses doubts: “whereas Millsap’s point, strictly speaking, may be true, it makes no practical difference when applied in real world settings and thus has been largely ignored.” However, it is important to note that, in the present context, the point is not that the requirement of measurement invariance should be raised to the level of a *sine qua non* for research in general—that would be wrong, because its importance clearly depends on pragmatic

dimensions of the research context, i.e., what the data are used for, as well as on the severity and nature of violations of measurement invariance, as Clark correctly indicates. The point is that Millsap's work shows *that prediction invariance cannot be adduced as evidence for measurement invariance*. This implies that the duty to investigate to what extent the inconsistency between measurement invariance and prediction invariance is important to real world settings lies *not* with those who think measurement invariance should be more central; *it lies with those who want to adduce prediction invariance as evidence that there is no test bias*. And it is in *this* perspective that I find it astounding that the official APA and SIOP testing standards and guidelines, which should light the researcher's path on such issues, uncritically embrace an invalid line of argument without issuing a serious qualification or warning on this point.

Psychometrics and Experimental Psychology

Psychometric theory has largely been developed in the context of individual differences research and correlational psychology. Heiser argues that the importance of psychometric theory has become limited, because contemporary psychology is predominantly experimentally oriented; and, according to Heiser, in experiments psychometric theory is not needed. Hence, current psychology has no need for psychometric modeling.

In my opinion, this diagnosis is incorrect. Experimental psychology can benefit enormously from psychometric input, as is demonstrated by some of the best recent work in the field. In fact, one of the examples I used comes directly from experimental social psychology, where Blanton, Jaccard, Gonzales, and Christie (2006) demonstrated the importance of psychometrics for the interpretation of the implicit association test. Other examples include Raijmakers, Dolan, and Molenaar (2001), who used finite mixture modeling to investigate discrimination learning; Visser, Raijmakers, and Molenaar (in press), who applied hidden Markov models to study sequence learning; and Wicherts, Dolan, and Hessen (2005), who used multigroup confirmatory factor analysis to investigate the origin of stereotype threat effects. That these studies were published in top journals suggests that, whatever Heiser may think of the Archipelago of experimental psychology, its inhabitants actually find the psychometric approach to their problems quite refreshing.

Heiser (2006, p. 459) also thinks that "the great thing of explanatory of independent variables is that you do not need to measure them." This may be a truism if one equates explanatory variables with experimental conditions, but that is not what researchers are typically interested in. Just like questionnaire users are not interested in items, but in the attributes that these items measure, so experimentalists are not interested in experimental conditions, but in the processes that these conditions manipulate. In fact, questions like "Do manipulations x and y affect the same process?" are quite common in experimental psychology, and such questions are amenable to psychometric treatment. It is, of course, true that the psychometric theory of experimental manipulation is not well developed, but this is precisely the sort of venue that psychometricians are excellent candidates for exploring. In conclusion, I would say that the experimental Archipelago that, according to Heiser, renders psychometric theory obsolete, is actually a significant growth market for creative and substantively interested psychometricians.

Conclusion

Many interesting and important questions in psychology are either of a psychometric nature, or have a strong psychometric component. With respect to such questions, psychometrics has much to offer to psychology, but has so far not realized its potential. The time is right for

this to change. Many versatile modeling approaches have been developed in the past decades; with the current computing power, using such approaches is quickly becoming a realistic option for researchers. This is illustrated by the fact that, in the past few years, several excellent psychometric modeling approaches to empirical problems have surfaced in various areas of psychology. Psychometric modeling is gaining momentum; and we should use this momentum to create as much educational and research opportunities as possible for researchers who are able and willing to take a psychometric approach to their problems. When psychological researchers see that psychometric modeling allows them to investigate problems that are otherwise inaccessible, that it is not a dead-end street but a source of interesting problems and ideas, that it need not be all that much more difficult to execute than the current default methods of analysis but yields more insight into measurement problems, I think that psychometrics may become firmly entrenched in large parts of psychology fairly quickly. With good tactical planning, no shots need be fired, and we may witness a velvet revolution in the next few years.

References

- Blanton, H., Jaccard, J., Gonzales, P.M., & Christic, C. (2006). Decoding the implicit association test: Implication for criterion prediction. *Journal of Experimental Social Psychology, 42*, 192–212.
- Borsboom, D., Mellenbergh, G.J., & Van Heerden, J. (2004). The concept of validity. *Psychology Review, 111*, 1061–1071.
- Clark, L.E. (2006). When a psychometric advance falls in the forest. *Psychometrika, 71*, 447–450.
- Ellis, J., & Junker, B.W. (1997). Tail-measurability in monotone latent variable models. *Psychometrika, 62*, 495–523.
- Heiser, W. J. (2006). Measurement without copper instruments and experiment without complete control. *Psychometrika, 71*, 457–461.
- Holland, P.W., & Hoskens, M. (2003). Classical test theory as a first-order item response theory: Application to true-score prediction from a possibly nonparallel test. *Psychometrika, 68*, 123–149.
- Kane, M. (2006). In praise of pluralism. A comment on Brosboom. *Psychometrika, 71*, 441–445.
- Millsap, R.E. (1997). Invariance in measurement and prediction: Their relationship in the single-factor case. *Psychological Methods, 2*, 248–260.
- Raijmakers, M.E.J., Dolan, C.V., & Molenaar, P.C.M. (2001). Finite mixture models for simple discrimination learning. *Memory & Cognition, 29*, 659–677.
- Sijtsma, K. (2006). Psychometrics in psychological research: Role model or partner in science? *Psychometrika, 71*, 451–455.
- Visser, I., Raijmakers, M.E.J., & Molenaar, P.C.M. (in press). Characterizing sequence knowledge using online measures and hidden Markov models. *Memory & Cognition*.
- Wicherts, J.M., Dolan, C.V., & Hessen, D.J. (2005). Stereotype threat and group differences in test performance: A question of measurement invariance. *Journal of Personality and Social Psychology, 89*, 696–716.

Manuscript received 19 JULY 2006

Final version received 19 JULY 2006

Published Online Date: 23 SEP 2006